NOVEL FUNGAL TAXA IN AN ALASKAN BOREAL FOREST: PHYLOGENETIC AFFINITIES, ECOLOGIES, AND RIBOSOMAL RNA SECONDARY STRUCTURES

By

Daniel Glass

RECOMMENDED:

Dr. Link E. Olson

Dr. Naoki Takebayashi

Dr. Lawrence Duffy

Dr. D. Lee Taylor
Advisory Committee Chair

Dr. Christa Mulder
Chair, Department of Biology and Wildlife

APPROVED:

Dr. Paul Layer
Dean, College of Natural Science and Mathematics

Dr. Lawrence Duffy, Dean of the Graduate School

Dec 6, 2011

Date

NOVEL FUNGAL TAXA IN AN ALASKAN BOREAL FOREST: PHYLOGENETIC AFFINITIES, ECOLOGIES, AND RIBOSOMAL RNA SECONDARY STRUCTURES

A

THESIS

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

By

Daniel Glass, B.S.

Fairbanks, Alaska

December 2011

Abstract

Phylogenetic analyses suggest that a novel DNA sequence (NS1) found in a boreal forest soil-clone library belongs to the fungal kingdom but does not fall unambiguously within any known class. In order to determine if NS1 codes for an authentic ribosomal RNA (rRNA) gene-copy, I modeled ribosomal RNA secondary structure for four gene regions. Such analyses have never been used on environmental ribosomal sequences before. It appears that NS1 does code for an authentic gene-copy and is not a biological or lab artifact

I also elucidated the habitat preferences, horizon preferences, and fine-scale spatial structure of NS1 using molecular methods. I determined that NS1 was associated with spruce and was found in both the organic and mineral soil horizons. It appears to have a clumped distribution on the scale of a few meters and its spatial distribution shows little inter-annual variability.

Together these findings suggest that NS1 does represent an authentic gene-copy and also shed light on the ecology of this putative taxon. I hope future efforts will expand our understanding of both its identity and function.

Table of Contents

Page

Signature Page……i

Title Page……ii

Abstract……iii

Table of Contents……iv

List of Figures……vi

List of Tables……vi

Introduction……1

Chapter 1 Evaluation of the authenticity of a highly novel environmental sequence from boreal forest soil using ribosomal RNA secondary structure modeling ……6

1.1 Abstract……6

1.2 Introduction……7

1.3 Methods……10

1.4 Results……17

1.5 Discussion……21

1.6 Conclusions……28

1.7 Acknowledgements……29

1.8 Literature Cited……30

1.9 Tables……37

1.10 Figures……38

Chapter 2 Habitat preferences, micro-spatial structure, and temporal persistence of a novel fungal taxon in Alaskan boreal forest soils……48

2.1 Abstract……48

2.2 Introduction……48

2.3 Methods……50

2.4 Results/Discussion……54

2.5 Conclusions……58
2.6 Literature Cited……59
2.7 Tables……61
2.8 Figure……62
Conclusions……63
Literature Cited……65
Appendix……66

List of Figures

Page

Figure 1.1. Map of ribosomal gene-regions……38
Figure 1.2. Phylogenetic reconstructions of the kingdom Eumycota……39
Figure 1.3. ITS1 secondary structure for NS1……42
Figure 1.4. Linear regression plot of ITS1 length vs. MFE……43
Figure 1.5. 5.8S secondary structures for NS1……44
Figure 1.6. ITS2 secondary structure for NS1……45
Figure 1.7. Linear regression plot of ITS2 length vs. MFE……46
Figure 1.8. LSU secondary structure for NS1……47
Figure 2.1. Distribution of NS1 across the riparian black spruce site FP5C…… 62
Figure A.1. Phylogeny of ascomycetes including novel taxa NS3 and NS4……71
Figure A.2. Phylogeny of basal fungi including NS1, NS2 and NS5……72

List of Tables

Table 1.1. Comparison of basal fungal sequences and NS1 to highly conserved 5.8S motifs……37
Table 1.2. Comparison of compensatory base changes……37
Table 1.3. Comparison of highly conserved motifs within the ITS2 gene-region……37
Table 2.1. Presence of NS1 in soil-extract PCR survey……61
Table 2.2. Origins of NS1 sequence variants……61
Table A.1. Primers designed to target novel taxa……67

Introduction

The fungal kingdom has long been shrouded in mystery. We still know much less about the kingdom Eumycota than we know about either the Plant or Animal kingdoms. This dearth of knowledge spans both our understanding of fungal systematics and our understanding of fungal ecology. The advent of microbe culturing increased our level of understanding of this kingdom but many fungi are difficult, or currently impossible, to grow in culture. Although the percentage of culturable fungi may not be as low as the corresponding proportion of bacteria (<1%; Pace, 1997) it is most likely no higher than 30% and more likely below 10% (Allen et al., 2003). Many fungal species are quite difficult to differentiate based on morphological characters alone. Furthermore, many potential fungal habitats have been poorly studied. Finally, having such a large number of fungal taxa (>1.5 million; Hawksworth, 2001; Blackwell, 2011) and such a small number of well-funded fungal taxonomists has led to a bottleneck for formally describing fungal species currently in collections. Because of these limitations far fewer fungal species have been formally described than are estimated to exist.

The number of fungal species has been conservatively estimated at 1.5 million species (Hawksworth, 2001). More recent estimates put this figure closer to 5.1 million species (Blackwell, 2011). Nonetheless, fewer than 100,000 species have been formally described (Blackwell, 2011). Considering the number of described species and the number of estimated species differ by over an order of magnitude there is great potential for new fungal taxa to be discovered. Because fungi perform prominent roles in several biogeochemical cycles through their roles as decomposers and plant mutualists, and provide direct services to humanity by providing us with many foods, medicines, and model organisms (Blackwell, 2011), the incentive to unearth this remaining diversity is growing.

During the last few decades, molecular techniques have surmounted many of the hurdles that constrained traditional mycological methods and have

revolutionized our understandings of fungal phylogenetics. Polymerase chain reaction (PCR)-based studies of soil DNA extracts have made it possible to directly study fungi in soils, without culturing. Such studies have uncovered a plethora of fungal diversity, including prominent novel lineages, greatly expanding our understanding of the evolutionary history of this kingdom. This molecular revolution not only restructured our understanding of fungal systematics, it also greatly benefited our understanding of fungal ecology. With the ability to study fungi directly *in situ* many questions that could best be addressed in a natural system were now feasible to study. Among many other insights, our understanding of fungal community composition, fungal diversity within various habitats, individual species distributions, individual species habitat preferences, and how ecological and biogeochemical factors impact the aforementioned phenomena grew substantially because of this molecular revolution.

Throughout this mycological renaissance, the loci of choice for both phylogenetic and ecological studies have been the ribosomal RNA (rRNA) gene-regions. However, no single gene-region is the perfect candidate for every DNA-based study. Each gene has its own history and this history may not perfectly match organismal history. The ribosomal RNA (rRNA) gene-region is the DNA locus most widely used by mycologists for both phylogenetic and ecological inquiries, which makes comparisons between studies possible. Ribosomes are found in all eukaryotes, hence rRNA, a component of ribosomes, and the genes that code for rRNA are present in all fungi. Because many other characters may be cryptic or absent in certain fungi, the universal rDNA locus makes comparisons across the fungal kingdom possible. The rDNA is abundant because it is coded in multiple tandem copies (Balakirev and Ayala, 2003). Therefore, rDNA should be easy to amplify even from rare taxa. Unlike protein-coding genes in which a single nucleotide mutation can cause dramatic changes to protein structure, particularly through the insertion of stop codons, RNA-coding genes can undergo a wider variety of changes and still produce a functional molecule. As long as regions that play an

essential role in rRNA processing remain unaltered, selection will not lead to the elimination of altered gene-copies. Consequently, both conserved regions useful for resolving deep phylogenetic relationships and more variable regions suitable for differentiating between species are present in rDNA. This makes it a prime candidate for reconstructing phylogenies at multiple levels and barcoding of fungal species.

In 2002, Alaskan fungal diversity appeared to be represented only by sporocarp (fruiting body) collections in mycological herbaria and a small number of fungal DNA sequences. These DNA sequences originated from fungal sporocarps collected in Southeast Alaska and a single study of ectomycorrhizal root tips collected in Southcentral Alaska (Lilleskov *et al.,* 2002). It appeared that no DNA-based study of fungal diversity in Alaskan boreal forest soils had been pursued nor had any prior DNA-based studies been pursued on Alaskan boreal fungi. Due to the limitations described above, DNA-based studies appeared to be the most promising way to estimate fungal diversity and determine community structure within Alaskan boreal forest soils. Soil cores representing all major stages of plant succession were collected from Bonanza Creek LTER (BNZ-LTER) sites, near Fairbanks, Alaska by Dr. Lee Taylor (University of Alaska Fairbanks) and colleagues (Taylor *et al.,* 2010). These soil cores were then separated by soil horizon. Soils are composed of several layers, or horizons, with different biogeochemical properties. They therefore likely harbor different microbial communities. The litter layer, which consists of plant matter that has not yet undergone noticeable decomposition, was discarded. The organic, humic, and uppermost mineral horizons were separated and saved. DNA was extracted from each horizon for each soil core and a DNA marker-region was targeted to selectively amplify fungi using PCR. By PCR-amplifying all of the fungi present in a soil extract it is possible to construct a census of the soil fungal community and determine the diversity and community composition of soil fungi. These fungal DNA amplicons were then cloned into *E. coli* cells. Only a single DNA fragment can be incorporated into a single cell, so each clone should represent a

single fungal sequence. Over 100,000 clones were sequenced. Through this process Taylor and colleagues were able to obtain a comprehensive estimate of fungal diversity and structure within the Bonanza Creek LTER (Taylor *et al.*, 2010).

The majority of these sequences appeared to fall within known fungal genera (D.L. Taylor, personal communication). At least one sequence, hereafter referred to as novel sequence one (NS1), appeared to be particularly divergent from known fungal phyla. If authentic, this sequence may represent a new class or even phylum of fungi. To put this in perspective, the Phylum Chordata includes all animals with a spinal chord while Reptilia and Mammalia are both classes within the Phylum Chordata.

Although the rRNA gene-regions are widely employed in molecular mycology, to my knowledge, no putatively fungal environmental sequences have had rRNA secondary structure modeled to confirm that a sequence is not a biological or lab artifact. Considering the high volume of environmental sequences being deposited in online molecular databases such as the International Nucleotide Sequence Databases (INSD), many of which may represent undescribed species, ensuring that only legitimate functional-gene copies are present is imperative. In Chapter 1 I describe the methods employed to evaluate the chances that NS1 represents an authentic gene-copy.

Because my results from Chapter 1 suggest that NS1 represents a real fungal taxon, and could potentially represent a highly novel fungal taxon, in Chapter 2 I show how molecular methods alone were used to gain preliminary insights into the ecology of this organism. I was able to determine the distribution of NS1 across LTER sites, its soil horizon preferences, and its spatial structure and temporal persistence within a site solely through selectively targeting DNA.

Lastly, in the Appendix, I show the fungal phylogenies containing four novel taxa detected in the same fashion as NS1 and report their possible phylogenetic affinities. The phylogenetic affinities of NS1 are mentioned in Chapter 1. In the Appendix I also describe the methods used in my efforts to amplify longer gene-

regions that could have further resolved the phylogenetic affinities of these taxa, including NS1, and strengthen the support for their placements in the fungal tree of life.

Chapter 1:

Evaluation of the authenticity of a highly novel environmental sequence from boreal forest soil using ribosomal RNA secondary structure modeling[1]

1.1 Abstract

Both the number of formally described species and the number of uncultured environmental DNA sequences deposited in the International Nucleotide Sequence Databases have increased substantially over the last two decades. Although the majority of these sequences represent authentic gene-copies, there is evidence of DNA artifacts in these databases as well. These include lab artifacts, such as chimeras, and biological artifacts such as pseudogenes. Sequences that fall in basal positions in phylogenetic trees and appear distant from known sequences are particularly likely to be artifacts. I have applied secondary structure analyses to determine if one such divergent sequence represents an authentic gene-copy. To my knowledge, such analyses have never been used on environmental ribosomal sequences.

Phylogenetic analyses suggest that a novel sequence (NS1) found in a boreal forest soil clone library belongs to the fungal kingdom but does not fall unambiguously within any known class. In order to determine if NS1 codes for an authentic ribosomal RNA (rRNA) gene, ribosomal secondary structure was modeled for four rRNA gene-regions (ITS1, 5.8S, ITS2, 5' LSU). These models were analyzed for the presence of conserved domains, conserved nucleotide motifs, and compensatory base changes. Minimal free energy (MFE) foldings and GC contents of sequences representing the major fungal clades, as well as NS1, were compared. Linear regression was used to determine if a relationship between ITS lengths and MFE was present. Together these findings suggest the inference that NS1 represents

---

[1] Glass, D.J., L.E. Olson, N. Takebayashi, D.L. Taylor. Prepared for submission to Molecular Phylogenetics and Evolution

an authentic gene-copy. The methods described here can be used on any rRNA-coding sequence, not just environmental fungal sequences. As second-generation sequencing methods that yield shorter sequences become more widely implemented, methods that evaluate sequence authenticity should also be more widely implemented. For fungi, the adjacent 5.8S and ITS2 gene-regions should be prioritized. This region is not only suited to distinguishing between closely related species, but it is also more informative regarding the authenticity of a sequence.

1.2 Introduction

The number of uncultured environmental fungal DNA sequences in nucleotide databases, such as the International Nucleotide Sequence Databases (INSD), has grown exponentially during the last two decades. For example, the number of newly-deposited fungal internal transcribed spacer (ITS) sequences in the INSD rose from roughly 2000 sequences in 2000 to roughly 22,000 in 2007 (Ryberg *et al.,* 2009). Sequences in the INSD are frequently used as reference sequences in phylogenetic reconstructions (Bridge *et al.,* 2003) or in BLAST (Altschul *et al.,* 1999) searches to determine sequence identity. Accurate phylogenetic reconstructions are based on comparisons of orthologous sequences, and if unidentified lab artifacts or biological artifacts such as pseudogenes are present, these phylogenetic inferences may be misleading (Bensasson *et al.,* 2001; Olson and Yoder, 2002). Therefore, it is imperative that sequences accessioned into the INSD do not represent unidentified biological or lab artifacts. Several types of artifactual rDNA have been encountered, such as chimeras (Osborne *et al.,* 2005) and pseudogenes (including numts; Anthony *et al.,* 2007).

Chimeras are common artifacts that can confound phylogenetic inferences. Chimeras can result from one of two separate phenomena. During PCR, when two different sequences are used as template DNA, the resulting product may include a combination of these original sequences (Jumpponen, 2007). A study by Jumpponen

(2007) found an alarmingly high proportion (>30%) of sequences in two of his clone libraries to be chimeric. Many of these chimeric sequences appeared to be more basal in the phylograms produced than the DNA sequences from which they were derived. If undetected, this could have led to the spurious inference of novel higher-level taxa (Hugenholtz and Huber, 2003). Post-PCR chimeric DNA sequences can also be generated when multiple short contiguous or minimally-overlapping PCR products from different template sequences are inadvertantly joined to produce longer continuous sequences, as can happen in studies of ancient or degraded DNA (Olson and Hassanin, 2002).

Pseudogenes are duplicated copies of a gene that originated from a functional gene-copy but are no longer functional. Therefore, they are no longer constrained by the same selective pressures as their functional counterparts (Perna and Kocher, 1996). In the case of eukaryotic tandemly-repeated nuclear ribosomal genes, this usually involves transposition to a novel chromosomal location, after which concerted evolution fails to maintain sequence homogeneity (Balakirev and Ayala, 2003). A particularly worrisome type of pseudogene for studies based on rDNA is the numt. Numts, or mitochondrial-derived nuclear pseudogenes, are the complementary genes in the mitochondrial genome that have been transposed into the nuclear genome (Richly and Leister, 2004). These are particularly difficult to detect, especially if recently derived, but are known to be misleading phylogenetic inferences (Perna and Kocher, 1996; Bensasson *et al.,* 2001; Olson and Yoder, 2002). How much impact the inclusion of a pseudogene in a phylogeny will have on phylogenetic inferences depends on both the age of the pseudogene and the degree of differentiation needed to adequately address the research questions being asked.

For example, the use of an anciently-derived pseudogene arising from a familiar, formally described taxon would most likely differ greatly in functionally-conserved regions of the sequence, and if detected, should be excluded from phylogenetic reconstructions. Otherwise, it could lead to long branches in phylograms that do not even remotely represent the true amount of evolutionary

distance between taxa. Alternatively, a recently-derived pseudogene representing a very novel and potentially anciently derived lineage would show little variation between paralogs compared to the amount of variation represented by the evolutionary history of the taxa being compared. If only deep phylogenetic affinities are being pursued and species-level resolution is not needed, the inclusion of a recently-derived pseudogene instead of the proper orthologous sequences should have little impact on tree topology.

The majority of studies of fungal phylogenetics, including mine, have targeted ribosomal RNA (rRNA)-coding sequences encompassing one or all of the following nuclear rDNA gene-regions: the small subunit (SSU), the large subunit (LSU), or the internal transcribed spacers (ITS) including the 5.8S region (Fig 1.1). Fortunately, rRNA-coding sequences need to preserve core rRNA secondary structures for proper RNA processing. This has led to the evolutionary conservation of certain domains and nucleotide motifs across the eukaryotic kingdom (Coleman, 2007). By analyzing the predicted secondary structure of an rRNA sequence and verifying that conserved domains and motifs are present, we can determine if the sequence is likely to code for functional rRNA (Harpke and Peterson, 2008; but see Olson and Yoder, 2002). To my knowledge, analyses of ribosomal secondary structure have never been undertaken on environmental sequences. This is surprising, given that analyzing secondary structure is one of the more widely-implemented tools for validating the authenticity of rRNA gene copies (Harpke and Peterson, 2008; Xiao *et al.,* 2010; Zheng *et al.,* 2008) and that community analyses of environmental microbial systems depend heavily on comparisons of only orthologous (versus paralogous) sequences.

A fungal clone library spanning partial ribosomal large subunit (LSU) and internal transcriber spacer (ITS) gene-regions constructed by Taylor *et al.* (2007) from boreal forest soils at the Bonanza Creek LTER near Fairbanks, AK, yielded a novel fungal DNA sequence (~1200 bp) that is divergent from known phyla and whose relationship to other fungi remains uncertain. Here, I employ various

methods to analyze the predicted secondary structure of this sequence to estimate the chance that it codes for an authentic rRNA gene copy. I also evaluate which rRNA gene-regions are most informative in this endeavor. Although the methods outlined have been employed on a fungal DNA sequence, they are applicable to any potentially novel rDNA sequence. I also attempt to determine the phylogenetic affinities of this sequence.

1.3 Methods

Site descriptions and sampling procedure

Soils used in this study originated from two Bonanza Creek LTER (BNZ-LTER, http://www.lter.uaf.edu) sites near Fairbanks, AK, USA. Soil cores were collected from BNZ-LTER site FP5C (64.71306662 °N, 148.1426888 °W) over the summer of 2003 (Taylor *et al.,* 2007). The site is a mature riparian stand on the left bank (southern terrace) of the Tanana River approximately 20 km southwest of Fairbanks. It is dominated by black spruce (*Picea mariana*) with a few tamaracks (*Larix laricina*). Soil characterizations by horizon have been described previously (Ping, 2000). Ninety-two soil cores were collected from nine 3 m^2 plots in a stratified random distribution across the 200 m by 200 m site. Each plot was divided into nine 1 m^2 subplots and soil cores were randomly taken from within each subplot. Soil cores were approximately 18 mm in diameter and 200 mm deep. The organic, humic, and mineral horizons were identified. Approximately 0.25 g of soil was taken as a subsample from each of these horizons, frozen at -80 °C, and then lyophilized.

During the summer of 2004, soil cores were collected from the upland (UP) BNZ-LTER sites. Fifty soil cores were collected at 10 m intervals along four parallel transects across each site. For each core, the organic and mineral soil horizons were separated and 1 g subsamples collected (Taylor *et al.,* 2010). In contrast to the

samples from site FP5C, subsamples were pooled by horizon before being frozen at -80 °C and lyophilized.

Initial DNA extraction, amplification, cloning, and sequencing

The molecular methods used to isolate and identify fungal DNA sequences from soils within both the riparian black spruce site and the upland site have been described in detail previously (Geml *et al.,* 2009; Geml *et al.,* 2010). In brief, the gene-region (~1200 bp) encompassing the ribosomal internal transcribed spacers (ITS) and a portion (~700 bp) of the ribosomal large subunit (LSU; Fig 1.1) was amplified from soil extracts using the fungal-specific PCR primers ITS1-F (Gardes and Bruns, 1993) and TW13 (Taylor and Bruns, 1999). Amplicons were cloned into pCR®4-TOPO vectors using a TOPO-TA kit (Invitrogen, Carlsbad, CA, USA) and sent to the Broad Institute of MIT and Harvard where transformations, automated clone-picking, and sequencing of clone libraries took place.

Identification of novel sequences within clone libraries

A BLAST (Altschul *et al.,* 1999) search against known fungi revealed a novel fungal DNA sequence within the black spruce site that is divergent from known phyla and whose relationship to other fungi is uncertain. Initial chimera checking was conducted by running separate BLAST sequence similarity searches for the ITS1 and ITS2 regions. For a more detailed description of these methods and stringency cutoffs, see Geml *et al.* (2009). This sequence was subsequently re-examined for chimerism using the program Uchime within the Uclust package (Edgar, 2010).

Primer design and reamplification of novel taxa

To evaluate the authenticity of this sequence, I designed primers to specifically target a 290 bp diagnostic region encompassing a portion of ITS1, the 5.8S, and a portion of ITS2 (Forward: L2F1 5'CCCGGTCGATATATTTACGAGAAG 3' and reverse: L2R2 5'GGGCAGAGATGAATATGCTAACAC 3'; Fig 1.1). Primer3 (Koressaar and Remm, 2007) was used to design primers. I used default settings except for the following changes: the 5.8S region was excluded as a potential priming site; minimum primer length was set to 21 bp, optimum to 24 bp, and maximum to 28 bp; minimum annealing temperature (primer Tm) was set to 55 °C, optimum was set to 62 °C, and maximum was set to 65 °C; and the optimal final product size was set to 300 bp. NetPrimer (http://www.premierbiosoft.com/netprimer/index.html) was used to cross-check the quality of potential primers. I verified that primer pairs had similar annealing temperatures and were not prone to self-dimerization or cross-dimerization. Finally, I used BLAST to verify that they would not amplify any known organisms besides the taxon I was targeting. Using these primers, I performed PCR on the same 2003 black spruce site organic DNA extracts from which NS1 was originally cloned.

Illustra PureTaq Ready-To-Go PCR beads (GE Healthcare, Buckinghamshire, UK) were used amplify NS1. The following program was run on an MJ Research PTC-225 Peltier Thermalcycler (Harlow Scientific, Arlington, MA, USA) for all PCR: an initial denaturation step of 96 °C for 2 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 40 s, elongation at 72 °C for 1 min, followed by a final elongation step of 72 °C for 10 min. Big Dye Terminator v3.1 (Applied Biosystems, CA, USA) chemistry was used for direct cycle sequencing. To purify BigDye Terminator reactions, I applied samples to 800 μL G-50 superfine Sephadex in Centri-Sep columns (Princeton Separations Inc., NJ, USA), followed by drying on a vacuum centrifuge. These products were sequenced on an ABI 3130xl Genetic

Analyzer (Life Technologies Corporation, Carlsbad, California, USA) at the University of Alaska Fairbanks Nucleic Acids Core Lab.

Phylogenetic analyses

In order to estimate the phylogenetic affinities of this novel taxon to other fungi, I added the ~700 bp 5' LSU region of the original NS1 sequence to an alignment constructed by Tim James, which was similar to that of White *et al.* (2006). The LSU is widely used in fungal phylogenetics because it is sufficiently conserved to construct multiple sequence alignments containing distantly-related taxa but variable enough to provide support for monophyletic genera and families. The 5.8S and SSU are more conserved and consequently do not offer as much fine-level resolution, while the ITS regions are hypervariable and consequently nearly impossible to use in multiple sequence alignments containing distantly related taxa. I constructed trees using parsimony, maximum likelihood, and Bayesian methods. The following programs were run on the UAF Life Science Informatics Portal: PAUP* v. 4.0b10 (Swofford, 2003), Modeltest 3.06 (Posada and Crandall, 1998), and MrModeltest (Nylander, 2004). In PAUP* I used the following settings to infer the optimal tree(s) under the parsimony criterion: a heuristic search with sequences added at random, tree bifurcation and reconstruction for branch swapping, and 1000 bootstrap replicates. Modeltest and MrModeltest were used to determine the best-fit evolutionary model (GTR+I+Γ) to be implemented in GARLI v. 1.0 (Zwickl, 2006) and MrBayes v. 3.1.2 (Posada and Buckley, 2004; Huelsenbeck and Ronquist, 2001) respectively.

The CIPRES portal (http://www.phylo.org/sub_sections/portal; Miller *et al.,* 2010) was used to run GARLI and MrBayes. I constructed phylogenetic trees using default parameters in GARLI v. 1.0 (Zwickl, 2006). Bootstrap support in GARLI was calculated based on 1000 iterations. In MrBayes, default settings were used besides the following: percentage burn-in= 0.25, generations = 1000000, sample frequency

= 1000, number of chains =4, and number of runs= 2. For all phylogenetic tree reconstructions nine outgroups representing basal ophisthokonts, but distinct from the kingdom Eumycota, were used. FigTree 1.31 (http://tree.bio.ed.ac.uk/software/figtree) was used to visualize and annotate trees.

Tests to evaluate the authenticity of NS1

I used a number of methods (described below) to evaluate functional attributes of the novel environmental sequence, NS1. Comparative analyses were carried out using sequences spanning the kingdom Eumycota from the Assembling the Fungal Tree of Life (AFTOL) project (http://aftol.org/). I chose these sequences for comparison because they were derived from vouchered and taxonomically identified herbarium specimens and isolates (Lutzoni *et al.,* 2004). Furthermore, the published sequences have been vetted using a series of automated and manual checks (Lutzoni *et al.,* 2004). Hence, of available sequences, these are highly likely, though not certain, to constitute authentic orthologs rather than pseudogenes or other artifacts. The vast majority of the sequences used in my phylogenetic reconstructions were also used in the analyses described below. Those that were excluded were omitted for one of two reasons: because I was unable to identify and separate gene-regions, particularly the SSU (which was not used in these analyses) or because a sequence comprising all four gene-regions was not present (Fig 1.1). GenBank sequences were used in place of AFTOL sequences if they exactly matched the corresponding AFTOL LSU sequence and, when possible, they also matched the corresponding sequence in the AFTOL 5.8S alignment (James *et al.*, 2006). GenBank sequences representing the species believed to have the shortest ITS gene-regions of any fungus were also modeled for comparison (Fujita *et al.,* 2001).

1) Modeling ribosomal secondary structure

I used a variety of programs from the Vienna RNA Websuite (http://rna.tbi.univie.ac.at/; Gruber *et al.,* 2008; Hofacker, 2003) to search for anomalous features within NS1 that might suggest an artifact. The secondary structure of NS1 was modeled using default settings in RNAfold. I modeled a folding of the entire original sequence, as well as independent foldings for each of the four constituent loci (ITS1, 5.8S, ITS2, partial LSU). Both minimum free energy structures and centroid structures were produced. For details on differences between MFE and centroid structures see Gruber *et al.* (2008). These were compared to previously modeled fungal secondary structures to determine if core conserved regions, including pan-eukaryotic homologies, were present.

RNAalifold (Bernhart *et al.,* 2008, Hofacker *et al.,* 2002) is a program designed to produce consensus secondary structures from multiple sequence alignments. Roughly 180 representative sequences of the major fungal lineages were extracted from the 5.8S and LSU AFTOL alignments of James *et al.* (2006) and used as a template against which to align my novel sequences by hand using SeAl 2.0 (http://tree.bio.ed.ac.uk/software/seal/). The LSU alignment predominately included the same taxa used in my phylogenetic reconstructions. These multiple sequence alignments were used to create consensus secondary structures of the 5.8S and 5' LSU regions using default settings in RNAalifold. This program also produces a multiple sequence alignment with the conserved stem regions from the consensus secondary structure and their positional probabilities (i.e., the chance that a given sequence has that base present at that position) superimposed onto the alignment. I was thus able to evaluate whether conserved regions in the primary sequence are also conserved in the secondary structure. I also compared output from RNAalifold to the RNAfold folding of the NS1 5.8S region to determine how many compensatory base changes were present in NS1.

In order to determine the minimal free energies (MFEs) of several sequences at once, I used the Quikfold program on the mfold web server (http://mfold.rna.albany.edu/?q=DINAMelt/Quickfold). Settings were changed to RNA, and the RNA 3.0 model for folding was used; all other settings were left at the defaults. One structure per sequence was produced. I performed linear regression in R v 2.8.1 (R Development Team, 2008) comparing locus length and MFE. The ITS1 regions of our reference sequences and the original NS1 sequence were used. The same analysis was performed separately on the ITS2 gene-region of these sequences. This let me estimate a typical MFE for a given sequence length and determine if NS1 fell within the 95% prediction interval for each locus.

2) Comparison of ITS1 and ITS2 length and GC contents

Prior studies have suggested that although length may differ between ITS1 and ITS2 for a given species it usually will have very similar GC contents for both gene-regions (Torres *et al.,* 1990; Harpke and Peterson, 2006; Zheng *et al.,* 2008; Mullineux and Hausner, 2009). ITS1 appears to have originated from an intergenic spacer (Clark, 1987) while ITS2 is derived from an expansion segment within the LSU (Nazar, 1980). Although the two regions originated independently, they are believed to undergo concerted evolution (Lalev and Nazar, 1999; Hausner and Wang, 2005, Mullineux and Hausner, 2009) and the retention of similar GC content is a vestige of this process (Torres *et al.,* 1990). Although theories about the biochemical evolutionary processes behind this phenomenon have been proposed, the underlying mechanism remains unknown (Torres *et al.,* 1990; Mullineux and Hausner, 2009). In general, because these regions are believed to undergo concerted evolution, pseudogenes frequently have higher AT content when released from this selective pressure (Harpke and Peterson, 2006). By combining GC analyses with secondary structure modeling it is also possible to determine if the base of conserved helices are GC rich. If the overall GC content of either gene-region

is particularly low, differs significantly between gene-regions, or if GC rich regions are absent, it is suggestive of a pseudogene (Zheng *et al.,* 2008).

Thus, pseudogenes or chimeras, neither of which are believed to undergo concerted evolution, might have divergent GC contents between ITS1 and ITS2, whereas functional genes should have similar GC contents because of selective constraints. To test this for NS1, the InfoSeq tool on the EMBOSS web server (Rice *et al.,* 2000) was used to calculate GC content of each locus. This same program was used to calculate the length of each locus.

In order to determine if NS1 fell within the range of natural variation of GC content for these loci, I compared values for NS1 to reference sequences obtained from the AFTOL website and GenBank. I determined the mean GC content and standard deviation for each locus separately, plotted the distribution of values, and determined where NS1 fell within this distribution. I then compared our results for ITS1 and ITS2 to determine if NS1 varied in GC content between loci.

1.4 Results

PCR and phylogenetic results

Using my taxon-specific primers I was able to amplify the 290 bp fragment of NS1 from several soil extracts besides the sample from which the original clone sequence was discovered. Sequences from these amplicons closely matched the original singleton clone. In addition, a ~1200 bp sequence very similar to NS1, differing at only eight sites, was found in a clone library originating from BNZ-LTER site UP2A, an upland mixed white spruce (*Picea glauca*) and birch (*Betula neoalaskana*) stand. This sequence not only originated from a different clone library that was constructed from a different DNA extract, it was amplified using primers slightly modified from those used to originally amplify NS1. I also reamplified the 290 bp marker for NS1 from soil extracts from this site. It is therefore nearly

impossible that this sequence arose from lab contamination by the original amplicon.

I was incapable of successfully amplifying and sequencing any portion of the LSU or SSU for NS1 using my original taxon-specific primers and any of the LSU or SSU primers I tested. I tested multiple new primers designed to be paired with a specific LSU or SSU primer and selectively amplify NS1. Even with these primers I could not reproduce portions of our original NS1 sequence longer than the 290 bp marker or amplify regions of the LSU beyond the length of my original NS1 amplicon.

Two most-parsimonious trees were obtained but both had the same tree topology and differed only in branch lengths for two sister taxa. The tree with the lowest likelihood score is shown. The maximum-likelihood tree corresponds to the single optimal tree under this criterion. The Bayesian tree shown is the consensus tree of two separate runs after their likelihoods converged.

Based on the phylogenies I produced from my original sequence and the similar sequence originating from the upland site it appears that NS1 has an uncertain placement in the kingdom Eumycota but nevertheless may belong to the basal fungal lineages (Fig 1.2a-c). This placement had practically no support (below 50% ML and parsimony bootstrap, and 67% posterior probability for smallest inclusive clade) and remains questionable (Fig 1.2a-c). Although NS1 is clearly fungal and appears to fall within the BFL, the lack of support for more precise placements suggests interpretations be treated with caution but also suggests that this may be because it is highly divergent from any known sequences.

A nucleotide discontiguous megaBLAST search found only one match for the ITS1 gene region of NS1. This uncultured soil sequence overlapped the entire sequence and shared 83% sequence identity (GenBank accession GQ921811.1). When only the ITS2 gene-region was screened, no significant matches were found. The ITS1-5.8S-ITS2 gene region shared the same top BLAST match with the ITS1 gene-region but produced many more results corresponding to partial matches for

the 5.8s gene-region. The entire NS1 sequence had top hits corresponding to environmental sequences with as high as 53% sequence overlap (corresponding predominately to 5.8S and LSU gene-regions) and maximum identities below 90%.

Secondary structure and GC analyses of NS1

ITS1

ITS1 secondary structure has not been surveyed in as much detail as that of ITS2 (Mullineux and Hausner, 2009). Nonetheless, among the fungi studied, this locus is characterized by having a main central hairpin and may have smaller hairpins branching off of this (Lalev and Nazar, 1998; Mullineux and Hausner, 2009). Even though this locus is relatively short in NS1 (90 bp), it possesses this main central hairpin (Fig 1.3). When ITS1 length was plotted against MFE, I found NS1 to fall within the 95% prediction interval (R^2=0.75; Fig 1.4). The GC content of the ITS1 gene-region for NS1 is within one standard deviation of the mean as well (n=155, mean=0.46, s.d.=0.12, NS1= 0.51).

5.8S

Similarities in 5.8S structure between the consensus folding and NS1 can be seen throughout the majority of the locus (Fig 1.5). Areas that do differ between NS1 and the consensus folding appear be more variable across fungi and are areas where NS1 displays similarities with sequences within the BFL.

Three regions within the 5.8S that have been proposed to be highly conserved across eukaryotes are motifs one, two, and three (Table 1.1; Harpke and Peterson, 2008). For NS1, motifs one and three are identical to both the sequence described in the literature and the consensus sequence. Motif two may be present in NS1, but the sequence does differ within this region (Table 1.1). However, this motif

differs in other fungi, particularly those belonging to the BFL as well. Of the 29 5.8S compensatory base changes (including GU pairs) shared across the fungal kingdom, I found NS1 to have 26 of these present (Table 1.2).

NS1 also fell within less than one-half a standard deviation of the mean MFE value of 5.8S foldings for these fungi (n=184, mean= -44.8, s.d.= 4.3, NS1= -46.9) and has a GC content within one standard deviation (n= 188, mean= 0.456, s.d.= 0.026, NS1= 0.43).

ITS2

An MFE folding of the NS1 ITS2 gene-region revealed the presence of structures conserved in most eukaryotes (Fig 1.6). The core structure of this region is a central bulge with four hairpin loops radiating from it. These are designated helices I-IV. Of these, helix II and III are the most conserved (Coleman, 2007). The ITS2 MFE folding for NS1 included helices II and III, but also had base pairings with low probability support within the central bulge region, and lacked the other two helices. Helix II is usually composed of fewer than 12 pairings, is never branched, and contains a U-U bulge at the base of the helix. NS1 did not have a U-U bulge at the base of helix II but did have this pair of bases present (Fig 1.6). Like many other fungi, NS1 had a bulge caused by another mispairing, in this case the presence of an unpaired guanine (Table 1.3). Helix III is the longest and contains a distinctive motif on the 5' end just before the apex (Schultz *et al.,* 2005; Wolf *et al.,* 2005; Coleman, 2007). This is most often UGGU but can differ slightly; variations such as UGG, GGU, and UGGGU have been observed (Schultz *et al.,* 2005). Helix III was present in NS1 and contained a motif similar, but not identical, to UGGU (Table 1.3, Fig 1.6).

When ITS2 length was plotted against MFE, I found NS1 to fall within the 95% prediction interval (R^2=0.51; Fig 1.7). The ITS2 GC content of NS1 was within one-half a standard deviation of the mean (n=155, mean=0.49, s.d.= 0.11, NS1= 0.52).

ITS1 vs. ITS2

NS1 had identical GC contents for both loci (51%). It also had very similar locus length (ITS1=90 bp; ITS2=97 bp).

LSU

The consensus folding for the LSU of all the reference fungi and NS1 had low probability support for many regions and is not shown. Foldings for two fungi, belonging to the BFL, are shown with NS1 for further comparison (Fig 1.8). Similarities in structure, especially near the 5' end, are noticeable. Many of these structures appear similar to those described for the B domain of the LSU (Ben Ali *et al.*, 1999; De Rijk *et al.,* 1999).

When compared to other fungal sequences for the conserved portion of the LSU gene-region I analyzed, NS1 had a predicted minimum free energy value for the folding that fell within one standard deviation of the mean (n= 182, mean= 93.2, s.d.= 9.2, NS1= 101.6). The GC content for this region of rRNA in NS1 fell within one standard deviation of the mean as well (n=175, mean= 0.465, s.d.= 0.035, NS1= 0.49).

1.5 Discussion

Phylogenetic and BLAST results suggested that NS1 is highly divergent from any previously described taxa. The low level of sequence similarity to known fungi across the 5' LSU suggests that this sequence might represent a novel class or even phylum of fungi. However, I was unable to amplify longer portions of the LSU or amplify the SSU for NS1. These additional characters might have assisted in more robustly-supported phylogenetic reconstructions.

Several factors could have contributed to my inability to reamplify DNA fragments for NS1 greater than 290 bp. The quality of DNA in the DNA extracts could have diminished over time due to multiple freeze-thaw cycles. This could have led to the sheering of longer DNA fragments, allowing short regions such as the 290 bp region I targeted to be amplified while longer fragments like those I tried to amplify for the LSU and SSU would no longer be intact. Due to NS1 being low in abundance (1 out of 384 passing clones in 2003 FP5C organic clone library; 1 out of 931 passing clones in 2004 UP2A organic clone library) there would be a high chance that PCR would be unsuccessful. For more dominant taxa, even though the same proportion of template has been sheered, because of the higher template quantity there is a higher chance that enough template DNA was present for successful PCR.

Additionally, when the primers I designed were paired with widely used SSU and LSU primers they appeared to be prone to dimerization. Since primer-dimer formation is a competitive reaction with template DNA amplification in PCR, this issue is more prominent when targeting taxa present at low concentrations such as NS1. Although PCR conditions were optimized to reduce this issue as much as possible, it may have led to my inability to amplify NS1 well enough to successfully clone or directly sequence these amplicons. Only a limited number of potential primers can be designed that are NS1-specific. Due to the small size of the more variable regions (ITS1 and ITS2) that could harbor potential priming sites an ideal primer could not be found. An ideal primer would be both taxon-specific and have little or no potential for primer-dimer formation with itself or the SSU and LSU primers with which it would be paired.

Evidence suggesting that NS1 is an authentic rRNA gene-copy

My results suggest that NS1 represents a functional rDNA gene copy from a potentially novel fungal taxon rather than a biological or lab artifact. Although there

is both evidence to support and evidence that calls into question the authenticity of the NS1 sequence, my results nonetheless support the view that it is a functional gene copy and thus truly represents an extremely novel fungal lineage.

If NS1 is a chimera, my secondary structure analyses would likely have provided an indication. Due to the nature of chimera formation and the necessity that certain structures be maintained for proper RNA processing and function, had NS1 been a chimera it would most likely have differed from other fungi in secondary structure far more than I detected. Furthermore, none of the chimera tests I conducted suggest that NS1 is a chimera. I also found a full length (1200 bp) sequence that is very similar to NS1, most likely representing the same lineage, in a separate clone library that was constructed using a different source DNA and amplified with different PCR primers. Finally, I was able to amplify a 290 bp region of NS1 from multiple different soil extracts. These results collectively suggest that NS1 is not simply a PCR artifact. The odds of creating the same chimera multiple times from different samples are nearly zero. Furthermore, NS1 appears to be particularly divergent from other fungi throughout the entire 1200 bp sequence, particularly the ITS regions. This also suggests that it is not a chimera. A chimera would be highly similar to other fungi throughout the portions of its sequence from which the templates originated.

The overall structure for all four loci was consistent with previous studies. Therefore, NS1 is not likely to be an ancient pseudogene. ITS1 secondary structure was similar to descriptions for *Schizosaccharomyces pombe* (Fig 1.3; Lalev and Nazar, 1998) and several other ascomycetes (Bridge *et al.,* 2008; Hausner and Wang, 2005; Mullineux and Hausner, 2009). It should be noted that all the fungal species that have had their ITS1 structure described are ascomycetes and that no basal fungi have had their ITS1 secondary structure modeled.

The overall structure of 5.8S for NS1 was similar to both that described in the literature and the 5.8S secondary structure I produced using the fungal consensus sequence (Fig 1.6; Vaughn and Sperbeck, 1984). NS1 also contained the two most

conserved hairpins within ITS2 (Fig 1.6). It appears that these hairpins are also the only ones present in ascomycetes with particularly short ITS2 gene regions (Hausner and Wang, 2005). Overall, these results suggest that NS1 has probably retained the most important regions of ITS2 for proper RNA processing.

Along with the overall secondary structures for all four loci being similar to previously described secondary structures, I also found that two of the most conserved domains within the 5.8S gene region, motifs one and three, were present in NS1. These domains play an important role in RNA processing and if absent or altered would strongly suggest a pseudogene (Table 1.1, Harpke and Peterson, 2008). Motif two did differ in NS1 from the sequence described by Harpke and Peterson (2008), but because this motif differed in other fungi as well, I believe it should not carry as much weight as the two other motifs (Table 1.1).

Comparing minimal free energies (MFE) can lend additional support to analyses of secondary structure (Harpke and Peterson, 2006; Zheng *et al.,* 2008). It quickly demonstrates if the sequence in question falls within the known range of fungal rRNA MFE values. I plotted both ITS1 and ITS2 length against MFE to determine if MFEs scale with length (Figs 1.4, 1.7). NS1 had MFEs consistent with the overall trend for structures of that length for both ITS1 and ITS2. This complements my findings from analyzing secondary structure and suggests that the most important structures were present, and the conformations for ITS1 and ITS2 were not due to chance.

Torres's (1990) study of ITS GC content only tested 20 organisms representing the entire eukaryotic domain. Although the study did include fungi, only the Dikarya were represented. My findings support the results of this study for most of the species tested, as have many other studies (Harpke and Peterson, 2006; Zheng *et al.,* 2008; Mullineux and Hausner, 2009; Xiao *et al.,* 2010). NS1 was particularly consistent with this rule, having identical GC content for both gene-regions.

Evidence calling into question the authenticity of NS1

The secondary structures and nucleotide motifs that were inconsistent with universal patterns described in the literature appeared to be exceptions in other organisms as well (Schultz *et al.,* 2005). As already mentioned, within the 5.8S region, the nucleotide sequence within motif two did differ from the universal sequence for eukaryotes slightly, but other basal fungal sequences used for comparison differed as well (Table 1.1). NS1 did have the highly conserved motifs one and three present while the basal fungi used for comparison showed variation within these regions as well (Table 1.1). These three motifs, particularly motif two, may not be as universally diagnostic for pseudogenes as suggested in Harpke and Peterson (2008), at least for basal fungi.

NS1 showed variation in conserved domains within ITS2 as well. The minimal free energy folding of ITS2 for NS1 did not include the UU bulge at the base of hairpin II that is widespread across the eukaryotic kingdom, but did include the two uracil bases. It is therefore reasonable to propose that the minimal free energy folding model did not accurately depict the *in vivo* conformation of this portion of RNA. Alternatively, NS1 as well as many other fungi appeared to have a bulge on helix two present but this was not always caused by unpaired uracils (Table 1.3). Perhaps the presence of a bulge is necessary for proper processing but the origin of the bulge is under less selective constraint.

NS1 does appear to have hairpin III, but the highly conserved UGGU motif near the 5' apex of the loop differed slightly, reading UGAU. Schultz *et al.* (2005) did mention that this motif is known to differ slightly in certain eukaryotes, although a UGAU variant was not mentioned. Because two of the regions most conserved across the entire eukaryotic kingdom did vary in our sequence, the authenticity of NS1 should be treated with some caution. However, these regions are known to vary in other organisms as well (Coleman, 2007) and are frequently absent or altered in basal fungi (Schultz *et al.,* 2005). I also found variation within these regions in

several of my reference sequences (Table 1.3). Therefore, they may not be as conserved in fungi as in other organisms.

Overall, the 5.8S and ITS2 regions were the most informative for comparing secondary structures and determining the presence of conserved regions. The secondary structure evidence that did not support NS1 representing functional rDNA also appeared to show variation in other fungal taxa. I therefore believe that this evidence should not carry as much weight as those regions that do not show variation across the fungal kingdom. Based on the weight of the evidence, I conclude that NS1 is most likely a functional gene-copy. NS1 may well be a pseudogene, but that would not seem to fully explain its divergence from all known fungal lineages.

Phylogenetic affinities

I set out to not only determine the secondary structure of NS1, but also to determine if it was fungal in origin. Our phylogenetic results suggest that it likely belongs to the Eumycota. Although unsupported by bootstrap and posterior probability values, it appears that NS1 could belong within the BFL.

Broader implications and future research

As microbial community ecology studies shift towards next-generation sequencing methods such as pyrosequencing, selecting a maximally informative region to amplify becomes more pressing, because less information is retained in shorter sequences (Nilsson *et al.,* 2009). If a locus will be used in phylogenetic reconstructions spanning the fungal kingdom, it is critical that the region targeted meet two criteria: 1) It should be useful for distinguishing between closely related species while also resolving deeper level relationships (Nilsson *et al.,* 2009); and 2) it should be useful for determining whether the sequence represents an authentic, potentially functional, gene-copy (Mai and Coleman, 1997; Schultz *et al.,* 2005).

When either the ITS1 or ITS2 gene-region is amplified with the flanking, highly-conserved, 5.8S gene-region, both the ITS1 and ITS2 gene-regions meet this first requirement, but ITS1 is less well suited to the latter.

The ITS1 gene-region has widely been proposed as the ideal marker because it varies more than ITS2 and is therefore better for distinguishing between closely related species (Chen *et al.,* 2001; Hinrikson *et al.,* 2005). One of the most often utilized forward primers in fungal ITS1 amplification is ITS1-F (Gardes and Bruns, 1993). A major issue with targeting ITS1 is the presence in many fungi of an intron at the 3' end of the SSU (Vralstad *et al.,* 2002; Perotto *et al.,* 2000). This intron is often amplified when the ITS1-F primer is used (Vralstad *et al.,* 2002), which may push the amplicon length beyond the reach of next-generation sequencing methods. The ITS2 gene-region shows moderate variation and can therefore still distinguish between species nearly as well in most cases and sometimes better than ITS1 (Nilsson *et al.,* 2008). However, ITS2 is better for determining whether rRNA secondary structure is maintained (Schultz *et al.,* 2005; Coleman, 2007). Not only are the conserved domains within ITS2 well described, there is a database containing these structures for thousands of organisms that can be used for comparison (Koetschan *et al.,* 2010). It therefore is logical to target this region preferentially over ITS1, even though it shows slightly less variation.

All of these analyses benefit from having a broad representation of fungi to be used for comparison. The basal fungal lineages have much poorer representation within the INSD than do the Dikarya. At the time of writing, over 2.3 million Dikarya nucleotide sequences had been deposited, while the BFLs were represented by fewer than 280,000. Of these roughly 280,000 sequences, approximately 268,000 belonged to the Neocallimastigomycota, while the Entomophoromycotina and Kickxellomycotina were both represented by fewer than 1000 nucleotide sequences (http://www.ncbi.nlm.nih.gov/taxonomy/txstat.cgi). Because the SSU has historically been preferentially targeted for the basal fungal lineages, this lack of

representation is particularly notable for the other ribosomal gene-regions (http://www.ncbi.nlm.nih.gov/taxonomy/txstat.cgi).

Recently, much more attention has been given to the chytrids. The outbreak of a worldwide amphibian decline due to a fungal pathogen, *Batrachochytrium dendrobatidis*, has prompted more research on this species (Rohr *et al.,* 2011; Voyles *et al.,* 2011). Nonetheless, complete ITS sequences for this species are not yet available in the INSD. Mycologists would benefit not only from continued research on this species, but all basal fungi. This is needed in order for unbiased genetic comparisons to be made and the most accurate phylogenies possible reconstructed.

If most known fungal species were represented and multiple sequences per species were available, it would be possible to determine nucleotide diversity (π) and identify nucleotide motifs conserved at the infraordinal level. These two methods, particularly the latter, have been shown to be the best way to identify pseudogenes in closely related species (Zheng *et al.,* 2008; Harpke and Peterson, 2006). Because close relatives to be used as reference sequences are needed for these methods to perform well, they were not applicable to my study at this time.

1.6 Conclusions

My phylogenetic results suggest that NS1 is fungal in origin. More precise phylogenetic placements might be made if longer gene regions could be amplified (Porter *et al.*, 2008). I modeled portions of the NS1 sequence to characterize ribosomal RNA secondary structure. Secondary structure analyses suggest that it represents an authentic gene-copy. MFE and GC analyses further supported the conclusion that it is not an artifact. Hence, NS1 may well be a pseudogene, but that would not seem to fully explain its divergence from all known fungal lineages. If shorter rRNA gene-regions are going to be targeted in the future, as next-generation sequencing becomes more widespread, our findings suggest prioritizing the 5.8S and ITS2 gene-regions. These regions are not only suited to distinguishing between

closely related species, but they are also more informative regarding the authenticity of a sequence.

1.7 Acknowledgements

We would like to thank Stefan Washietl for his valuable input on the applicability of different programs on the RNA Websuite to our research questions. We would also like to thank Tim James for supplying reference sequences, input, and assistance constructing phylogenies of the basal fungal lineages. The entire Taylor lab provided input and assistance throughout this entire research project and deserves thanks as well. This material is based on work supported by the National Science Foundation through awards EF-0333308 and DEB-0620579, Alaska EPSCoR NSF award #EPS-0701898 and the State of Alaska, and by the USDA Forest Service, Pacific Northwest Research Station through agreement number RJVA-PNW-01-JV-11261952-231.

1.8 Literature Cited

Altschul S.F., W. Gish, W. Miller, and E.W. Myers (1999). "Basic local alignment search tool." Journal of Molecular Biology **215:** 403-410.

Anthony, N.M., S.L. Clifford, M. Bawe-Johnson, K.A. Abernethy, M.W. Bruford, E.J. Wickings (2007). "Distinguishing gorilla mitochondrial sequences from nuclear integrations and PCR recombinants: Guidelines for their diagnosis in complex sequence databases." Molecular Phylogenetics and Evolution **43**: 553-566.

Balakirev, E. S. and F.J. Ayala (2003). "Pseudogenes: Are they "Junk" or functional DNA?" Annual Review of Genetics **37**: 123-151.

Ben Ali, A., J. Wuyts, R. D. Wachter, A. Meyer, and Y. Van de Peer (1999). "Construction of a variability map for eukaryotic large subunit ribosomal RNA." Nucleic Acids Research **27**: 2825-2831.

Bensasson, D., D. Zhang, D.L. Hartl, and G.M. Hewitt (2001). "Mitochondrial pseudogenes: evolution's misplaced witnesses." Trends in Ecology and Evolution **16:** 314-321.

Bernhart, S.H., I.L. Hofacker, S. Will, A.R. Gruber, and P.F. Stadler (2008). "RNAalifold: improved consensus structure prediction for RNA alignments." BMC Bioinformatics **9**: 13.

Bridge, P.D., P.J. Roberts, B.M. Spooner and G. Panchal (2003) "On the unreliability of published DNA sequences." New Phytologist **160**: 43-48.

Bridge, P.D., T. Schlitt, P.F. Cannon, A.G. Buddie, M. Baker, and A.M. Borman (2008). "Domain II hairpin structure in ITS1 sequences as an aid in differentiating recently evolved animal and plant pathogenic fungi." Mycopathologia **166**: 1-16.

Chen, Y.C., J.D. Eisner, M.M. Kattar, S.L. Rassoulian-Barrett, K. Lafe, U. Bui, A.P. Limaye, and B.T. Cookson (2001). "Polymorphic internal transcribed spacer region 1 DNA sequences identify medically important yeasts." Journal of Clinical Microbiology **39**: 4042-4051.

Clark, C.G. (1987). "On the evolution of ribosomal-RNA." Journal of Molecular Evolution **25**: 343-350.

Coleman, A.W. (2007). "Pan-eukaryote ITS2 homologies revealed by RNA secondary structure." Nucleic Acids Research **35**: 3322-3329.

De Rijk, P., E. Robbrecht, S. de Hoog, A. Caers, Y. Van de Peer, R. De Wachter (1999). "Database on the structure of large subunit ribosomal RNA." Nucleic Acids Research **27**: 174-178.

Edgar, R.C. (2010). "Search and clustering orders of magnitude faster than BLAST." Bioinformatics **26**: 2460-2461.

Fujita, S.I., Y. Senda, S. Nakaguchi, and T. Hashimoto (2001) "Multiplex PCR using internal transcribed spacer 1 and 2 regions for rapid detection and identification of yeast strains." Journal of Clinical Microbiology **39:** 3617 -3622.

Gardes, M. and T.D. Bruns (1993). "ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts." Molecular Ecology **2**: 113-118.

Geml, J., G.A. Laursen, I. Timling, J.M. McFarland, M.G. Booth, N. Lennon, C. Nusbaum, and D.L. Taylor (2009). "Molecular phylogenetic biodiversity assessment of arctic and boreal ectomycorrhizal *Lactarius* Pers. (Russulales; Basidiomycota) in Alaska, based on soil and sporocarp DNA." Molecular Ecology **18**: 2213-2227.

Geml, J., G.A. Laursen, I.C. Herriott, J.M. McFarland, M.G. Booth, N. Lennon, H.C. Nusbaum, and D.L. Taylor (2010). "Phylogenetic and ecological analyses of soil and sporocarp DNA sequences reveal high diversity and strong habitat partitioning in the boreal ectomycorrhizal genus *Russula* (Russulales; Basidiomycota)." New Phytologist **187**: 494-507.

Gruber, A.R., R. Lorenz, S.H. Bernhart, R. Neubock, and I.L. Hofacker (2008). "The Vienna RNA Websuite." Nucleic Acids Research **36**: W70-W74.

Harpke, D. and A. Peterson (2006). "Non-concerted ITS evolution in *Mammillaria* (Cactaceae)." Molecular Phylogenetics and Evolution **41**: 579-593.

Harpke, D. and A. Peterson (2008). "5.8S motifs for the identification of pseudogenic ITS regions." Botany **86**: 300-305.

Hausner, G. and X. Wang (2005). "Unusual compact rDNA gene arrangements within some members of the Ascomycota: evidence for molecular co-evolution between ITS1 and ITS2." Genome **48**: 648-660.

Hinrikson, H.P., S.F. Hurst, T.J. Lott, D.W. Warnock, and C.J. Morrisson (2005). "Assessment of ribosomal large-subunit D1-D2, internal transcribed spacer 1, and internal transcribed spacer 2 regions as targets for molecular identification of medically important *Aspergillus* species." Journal of Clinical Microbiology **43**: 2092-2103.

Hofacker, I.L., M. Fekete, P.F. Stadler (2002). "Secondary structure prediction for aligned RNA sequences." Journal of Molecular Biology **319**: 1059-1066.

Hofacker, I.L. (2003). "Vienna RNA secondary structure server." Nucleic Acids Research **31**: 3429-3431.

Huelsenbeck, J.P. and F. Ronquist. (2001). "MRBAYES: Bayesian inference of phylogeny." Bioinformatics **17**:754-755.

Hugenholtz, P. and T. Huber (2003). "Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases." International Journal of Systematic and Evolutionary Microbiology **53**: 289-293.

James, T.Y., F. Kauff, C.L. Schoch, P. B. Matheny, V. Hofstetter, C.J. Cox, G. Celio, C. Gueidan, E. Fraker, J. Miadlikowska, H.T. Lumbsch, A. Rauhut, V. Reeb, A.E. Arnold, A. Amtoft, J.E. Stajich, K. Hosaka, G. Sung, D. Johnson, B. O'Rourke, M. Crockett, M. Binder, J.M. Curtis, J.C. Slot, Z. Wang, A.W. Wilson, A. Schußler, J.E. Longcore, K. O'Donnell, S. Mozley-Standridge, D. Porter, P.M. Letcher, M. J. Powell, J.W. Taylor, M.M. White, G.W. Griffith, D.R. Davies, R.A. Humber, J.B. Morton, J. Sugiyama, A.Y. Rossman, J.D. Rogers, D.H. Pfister, D. Hewitt, K. Hansen, S. Hambleton, R. A. Shoemaker, J. Kohlmeyer, B. Volkmann Kohlmeyer, R.A. Spotts, M. Serdani, P.W. Crous, K.W. Hughes, K. Matsuura, E. Langer, G. Langer, W.A. Untereiner, R. Lucking, B. Budel, D.M. Geiser, A. Aptroot, P. Diederich, I. Schmitt, M. Schultz, R. Yahr, D.S. Hibbett, F. Lutzoni, D.J. McLaughlin, J.W. Spatafora and R. Vilgalys (2006). "Reconstructing the early evolution of Fungi using a six-gene phylogeny." Nature **443**: 818-822.

Jumpponen, A. (2007). "Soil fungal communities underneath willow canopies on a primary successional glacier forefront: rDNA sequence results can be affected by primer selection and chimeric data." Microbial Ecology **53**: 233-246.

Koetschan, C., F. Forster, A. Keller, T. Schleicher, B. Ruderisch, R. Schwarz, T. Muller,

M. Wolf, and J. Schultz (2010). "The ITS2 Database III-sequences and structures for phylogeny." Nucleic Acids Research **38**: D275-D279.

Koressaar, T. and M. Remm (2007). "Enhancements and modifications of primer design program Primer3." Bioinformatics **23**: 1289-1291.

Lalev, A.I. and R.N. Nazar (1998). "Conserved core structure in the internal transcribed spacer 1 of the *Schizosaccharomyces pombe* precursor ribosomal RNA." Journal of Molecular Biology **284**: 1341-1351.

Lalev, A.I. and R.N. Nazar (1999). "Structural equivalence in the transcribed spacers of pre-rRNA transcripts in *Schizosaccharomyces pombe*." Nucleic Acids Research **27**: 3071-3078.

Lutzoni, F., F. Kauff, C.J. Cox, D. McLaughlin, G. Celio, B. Dentinger, M. Padamsee, D. Hibbett, T.Y. James, E. Baloch, M. Grube, V. Reeb, V. Hofstetter, C. Schoch, A.E. Arnold, J. Miadlikowska, J. Spatafora, D. Johnson, S. Hambleton, M. Crockett, R. Shoemaker, G. Sung, R. Lucking, T. Lumbsch, K. O'Donnell, M. Binder, P. Diederich, D. Ertz, C. Gueidan, K. Hansen, R.C. Harris, K. Hosaka, Y. Lim, B. Matheny, H. Nishida, D. Pfister, J. Rogers, A. Rossman, I. Schmitt, H. Sipman, J. Stone, J. Sugiyama, R. Yahr, and R. Vilgalys (2004). "Assembling the fungal tree of life: Progress, classification and evolution of subcellular traits." American Journal of Botany **91**: 1446-1480.

Mai, J.C. and A.W. Coleman (1997). "The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants." Journal of Molecular Evolution **44**: 258-271.

Miller, M.A., W. Pfeiffer, and T. Schwartz (2010). "Creating the CIPRES Science Gateway for inference of large phlogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1-8. URL:http://www.phylo.org/sub_sections/portal. Accessed 2011-06-16.

Mullineux, T. and G. Hausner (2009). "Evolution of rDNA ITS1 and ITS2 sequences and RNA secondary structures within members of the fungal genera *Grosmannia* and *Leptographium*." Fungal Genetics and Biology **46**: 855-867.

Nazar, R.N. (1980). "A 5.8-S rRNA-like sequence in prokaryotic 23-S ribosomal-RNA." FEBS Letters **119**: 212-214.

Nilsson, R.H., E. Kristiansson, M. Ryberg, N. Hallenberg and K. Larsson (2008). "Intraspecific ITS variability in the kingdom Fungi as expressed in the international sequence databases and its implications for molecular species

identification." Evolutionary Bioinformatics **4**: 193-201.

Nilsson, R.H., M. Ryberg, K. Aberenkov, E. Sjokvist, and E. Kristiansson (2009). "The ITS region as a target for characterization of fungal communities using emerging sequencing technologies." FEMS Microbiology Letters **296**: 97-101.

Nylander, J.A.A. 2004. MrModeltest v2. Program distributed by the author. Uppsala, Sweden: Evolutionary Biology Centre, Uppsala University.

Olson, L.E. and A. Hassanin (2002). "Contamination and chimerism are perpetuating the legend of the snake-eating cow with twisted horns (*Pseudonovibos spiralis*). A case study of the pitfalls of ancient DNA." Molecular Phylogenetics and Evolution **27**: 545-548.

Olson, L.E. and A.D. Yoder (2002). "Using secondary structure to identify ribosomal numts: Cautionary examples from the human genome." Molecular Biology and Evolution **19**: 93-100.

Osborne C.A., M. Galic, P. Sangwan, P.H. Janssen (2005). "PCR-generated artefact from 16S rRNA gene-specific primers." FEMS Microbiology Letters **248**: 183-187.

Perna, N.T. and T.D. Kocher (1996). "Mitochondrial DNA: Molecular fossils in the nucleus." Current Biology **6**: 128-129.

Perotto, S., P. Nepote-Fus, L. Saletta, C. Bandi, and J.P.W. Young (2000). "A diverse population of introns in the nuclear ribosomal genes off ericoid mycorrhizal fungi includes elements with sequence similarity to endonuclease-coding genes." Molecular Biology and Evolution **17**: 44-59.

Ping, C. (2000). "Soil horizon descriptions/classification and lab analysis, Bonanza Creek LTER - University of Alaska Fairbanks." BNZ:149, URL http://www.lter.uaf.edu/data_detail.cfm?datafile_pkey=149

Porter, T.M., C.W. Schadt, L. Rizvi, A.P. Martin, S.K. Schmidt, L. Scott-Denton, R. Vilgalys, and J.M. Moncalvo (2008). "Widespread occurrence and phylogenetic placement of a soil clone group adds a prominent new branch to the fungal tree of life." Molecular Phylogenetics and Evolution **46**: 635-644.

Posada, D. and K.A. Crandall (1998). "MODELTEST: testing the model of DNA substitution." Bioinformatics **14**: 817-818.

Posada D. and T.R. Buckley (2004). "Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests." Systematic Biology **53**: 793-808

R Development Core Team (2008). "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria. ISBN 3 -900051-07-0, URL http://www.R-project.org.

Rice, P., I. Longden, and A. Bleasby (2000). "EMBOSS: The European molecular biology open software suite." Trends in Genetics **16**: 276-277.

Richly, E. and D. Leister (2004). "NUMTs in sequenced eukaryotic genomes." Molecular Biology and Evolution **21**: 1081-1084.

Rohr, J.R., N.T. Halstead, and T.R. Raffel (2011). "Modeling the future distribution of the amphibian chytrid fungus: the influence of climate and human-associated factors." Journal of Applied Ecology **48**: 174-176.

Ryberg, M., E. Kristiansson, E. Sjokvist, and R. H. Nilsson (2009). "An outlook on the fungal internal transcribed spacer sequences in GenBank and the introduction of a web-based tool for the exploration of fungal diversity." New Phytologist **181**: 471-477.

Schultz, J., S. Maisel, D. Gerlach, T. Muller, M. Wolf (2005). "A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota." RNA-a Publication of the RNA Society **11**: 361-364.

Swofford, D.L. (2003). "PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)." Version 4. Sinauer Associates, Sunderland, Massachusetts.

Taylor, D.L. and T.D. Bruns (1999). "Community structure of ectomycorrhizal fungi in a *Pinus muricata* forest: minimal overlap between the mature forest and resistant propagule communities." Molecular Ecology **8**: 1837-1850.

Taylor, D.L., I.C. Herriott, J. Long, and K. O'Neill (2007). "TOPO TA is A-OK: a test of phylogenetic bias in fungal environmental clone library construction." Environmental Microbiology **9**: 1329-1334.

Taylor, D.L., I.C. Herriott, K.E. Stone, J.W. McFarland, M.G. Booth, and M.B. Leigh (2010). "Structure and resilience of fungal communities in Alaskan boreal forest soils." Canadian Journal of Forest Research **40**: 1288-1301.

Torres, R.A., M. Ganal, and V. Hemleben (1990). "GC balance in the internal

transcribed spacers ITS-1 and ITS-2 of nuclear ribosomal-RNA genes." Journal of Molecular Evolution **30**: 170-181.

Vaughn, J.C., and S.J. Sperbeck (1984). "A universal model for the secondary structure of 5.8S ribosomal-RNA molecules, their contact sites with 28S ribosomal-RNAs, and their prokaryotic equivalent." Nucleic Acids Research **12**: 7479-7502.

Voyles, J., E.B. Rosenblum, and L. Berger (2011). "Interactions between *Batrachochytrium dendrobatidis* and its amphibian hosts: a review of pathogenesis and immunity." Microbes and Infection **13**: 25-32.

Vralstad, T., E. Myhre, and T. Schumacher (2002). "Molecular diversity and phylogenetic affinities of symbiotic root-associated ascomycetes of the Helotiales in burnt and metal polluted habitats." New Phytologist **155**: 131-148.

White, M.M., T.Y. James, K. O'Donnell, M.J. Cafaro, Y. Tanabe, and J. Sugiyama (2006). "Phylogeny of the Zygomycota based on nuclear ribosomal sequence data." Mycologia **98**: 872-884.

Wolf, M., M. Achtziger, J. Schultz, T. Dandeker, and T. Muller (2005). "Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures." RNA-a Publication of the RNA Society **11**: 1616-1623.

Xiao, L.Q., M. Moller, and H. Zhu (2010). "High nrDNA ITS polymorphism in the ancient extant seed plant *Cycas*: Incomplete concerted evolution and the origin of pseudogenes." Molecular Phylogenetics and Evolution **55**: 168-177.

Zheng, X.Y., D.Y. Cai, L. Yao, and Y. Teng (2008). "Non-concerted ITS evolution, early origin and phylogenetic utility of ITS pseudogenes in *Pyrus*." Molecular Phylogenetics and Evolution **48**: 892-903.

Zwickl, D.J. (2006). "Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion." Ph.D. dissertation, The University of Texas at Austin.

1.9 Tables

Table 1.1. Comparison of basal fungal sequences and NS1 to highly conserved 5.8S motifs described in Harpke and Peterson (2008).

	Motif 1	Motif 2	Motif 3
Consensus (Literature)	CGATGAAGAACGCAGC	GAATTGCAGAATTC	TTTGAACGCA
NS1	CGATGAAGAACGCAGC	GA***T***TTGCA-A***C***TTC	TTTGAACGCA
Conidiobolus coronatus	CGATGAAGAACG***TT***GC	GAATTGCAG***TC***TT***T***	TTTGAAC***C***CA
Batrachochytrium dendrobatidis	CGATGAAGAACGCAGC	GAATTGCAG-AA***CCT***	TTTGAACGCA

Table 1.2. Comparison of compensatory base changes (CBCs) and non-canonical base pairs (e.g., GU pairs) for the 5.8s consensus sequence produced using RNAalifold and the 5.8s sequence produced using RNAfold.

	CBCs	GU pairs	Total
Consensus	26	3	29
NS1	20	6	26

Table 1.3. Comparison of highly conserved motifs within the ITS2 gene-region. Fungi representing most of the major clades are compared to the consensus motif described in previous literature. Asterisks denote that the group given is undergoing taxonomic revision and is not a true phylum.

Taxon	Phylum	Hairpin II Bulge	Hairpin III 5' Motif
Consensus (Literature)	NA	UU Mismatch	UGGU
NS1	NA	Unpaired G	UGAU
Basidiobolus ranarum	BFL*	Unpaired G	UGGU
Conidiobolus coronatus	BFL*	Unpaired U	AGUU
Neocallimastix frontalis	BFL*	UU Mismatch	UGGU
Sphaeronaemella fimicola	Ascomycota	Unpaired A	AGU
Taphrina wiesneri	Ascomycota	UU Mismatch	AGGU
Phanaerochaete chyrsosporium	Basidiomycota	UU Mismatch	UGGU

1.10 Figures

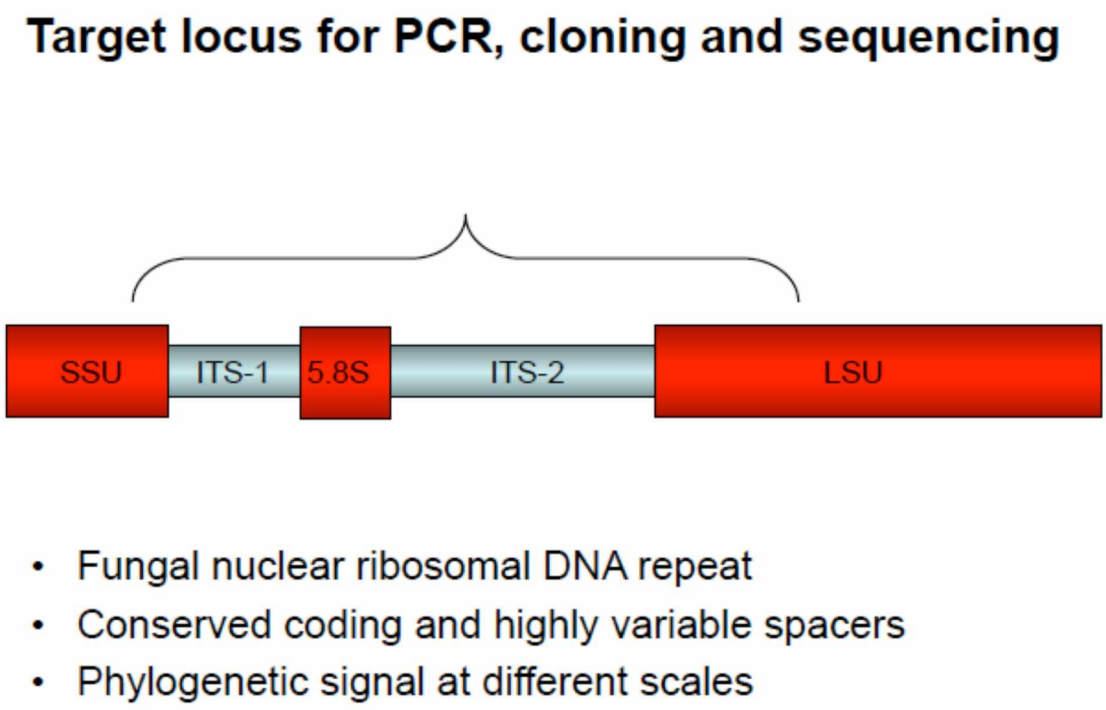


Figure 1.1. Map of ribosomal gene-regions.

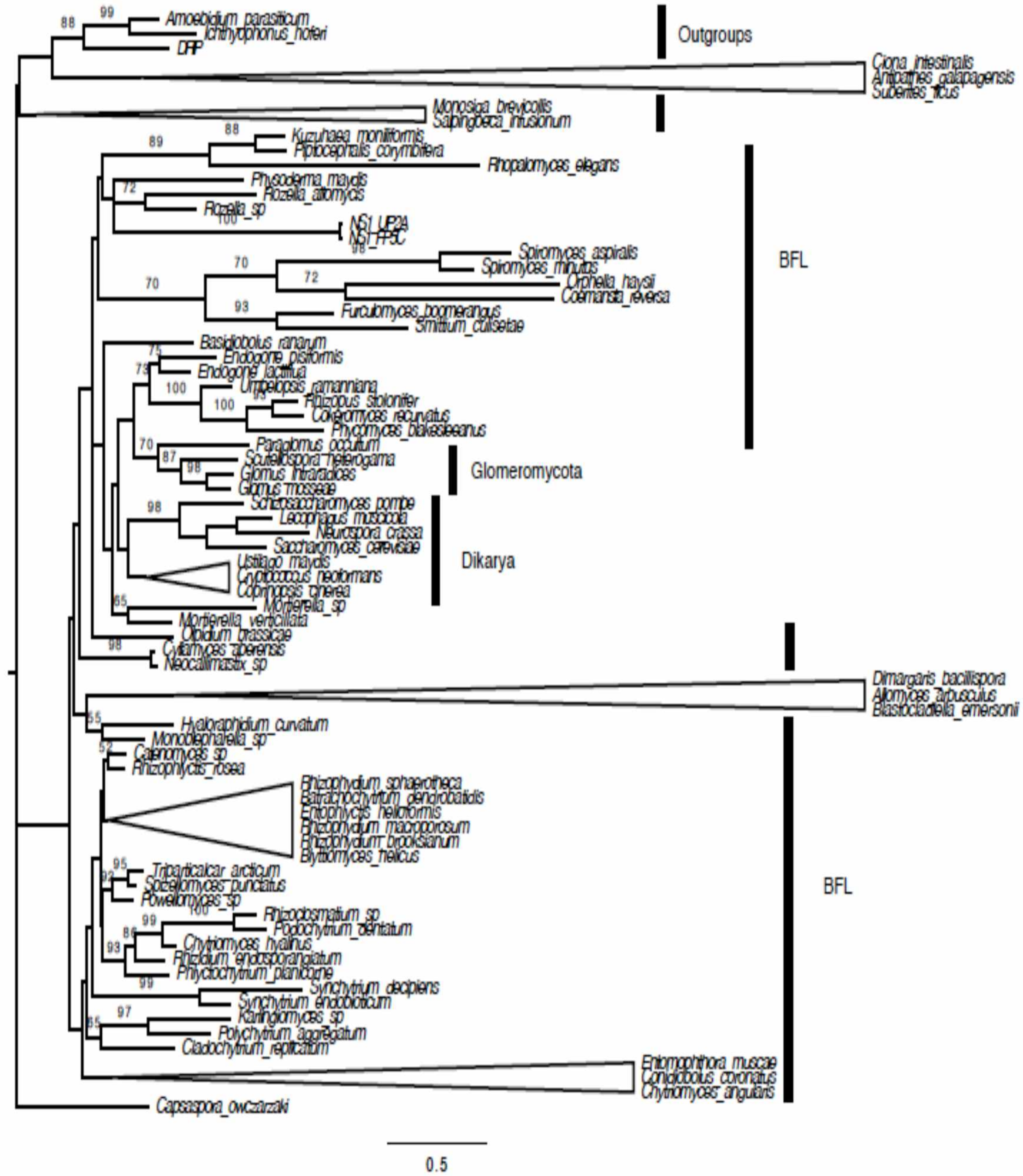


Figure 1.2a. Phylogenetic reconstructions of the kingdom Eumycota based on partial LSU rDNA gene-region. Includes NS1 sequences from both the riparian black spruce site (FP5C) and upland site (UP2A). Numbers indicate bootstrap support. If support for a clade (besides the smallest containing NS1) had below 50% bootstrap support the clade was collapsed. Based on alignment by James. **a) Parsimony tree constructed using PAUP*.**

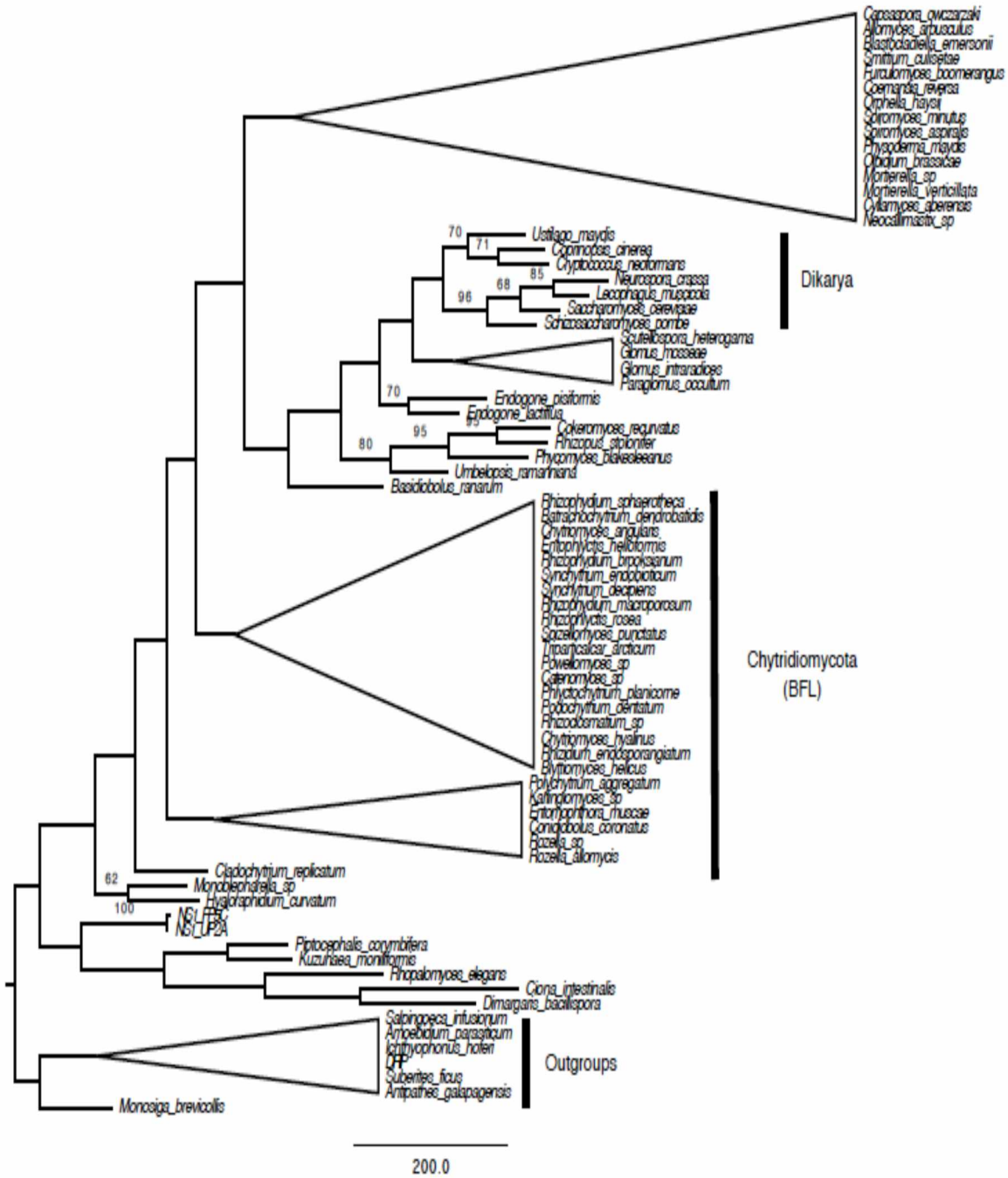


Figure 1.2b. Phylogenetic reconstructions of the kingdom Eumycota based on partial LSU rDNA gene-region. Includes NS1 sequences from both the riparian black spruce site (FP5C) and upland site (UP2A). Numbers indicate bootstrap support. If support for a clade (besides the smallest containing NS1) had below 50% bootstrap support the clade was collapsed. Based on alignment by James. **b) Maximum likelihood tree constructed using GARLI.**

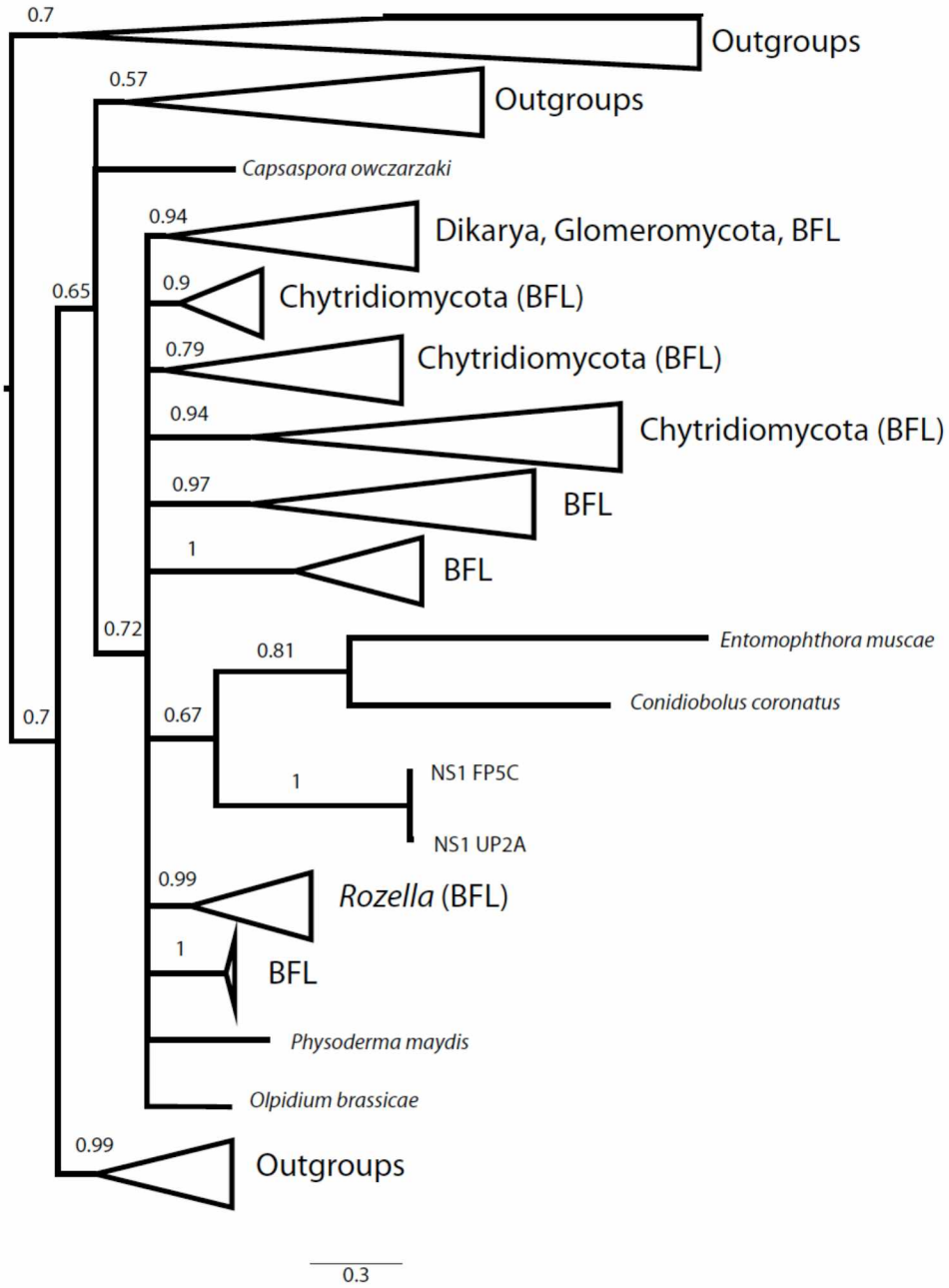


Figure 1.2c. Phylogenetic reconstructions of the kingdom Eumycota based on partial LSU rDNA gene-region. Includes NS1 sequences from both the riparian black spruce site (FP5C) and upland site (UP2A). Numbers indicate posterior probabilities. If support for a clade (besides the smallest containing NS1) had below 90% posterior probability support the clade was collapsed. Based on alignment by James. **c) Bayesian tree constructed using MrBayes.**

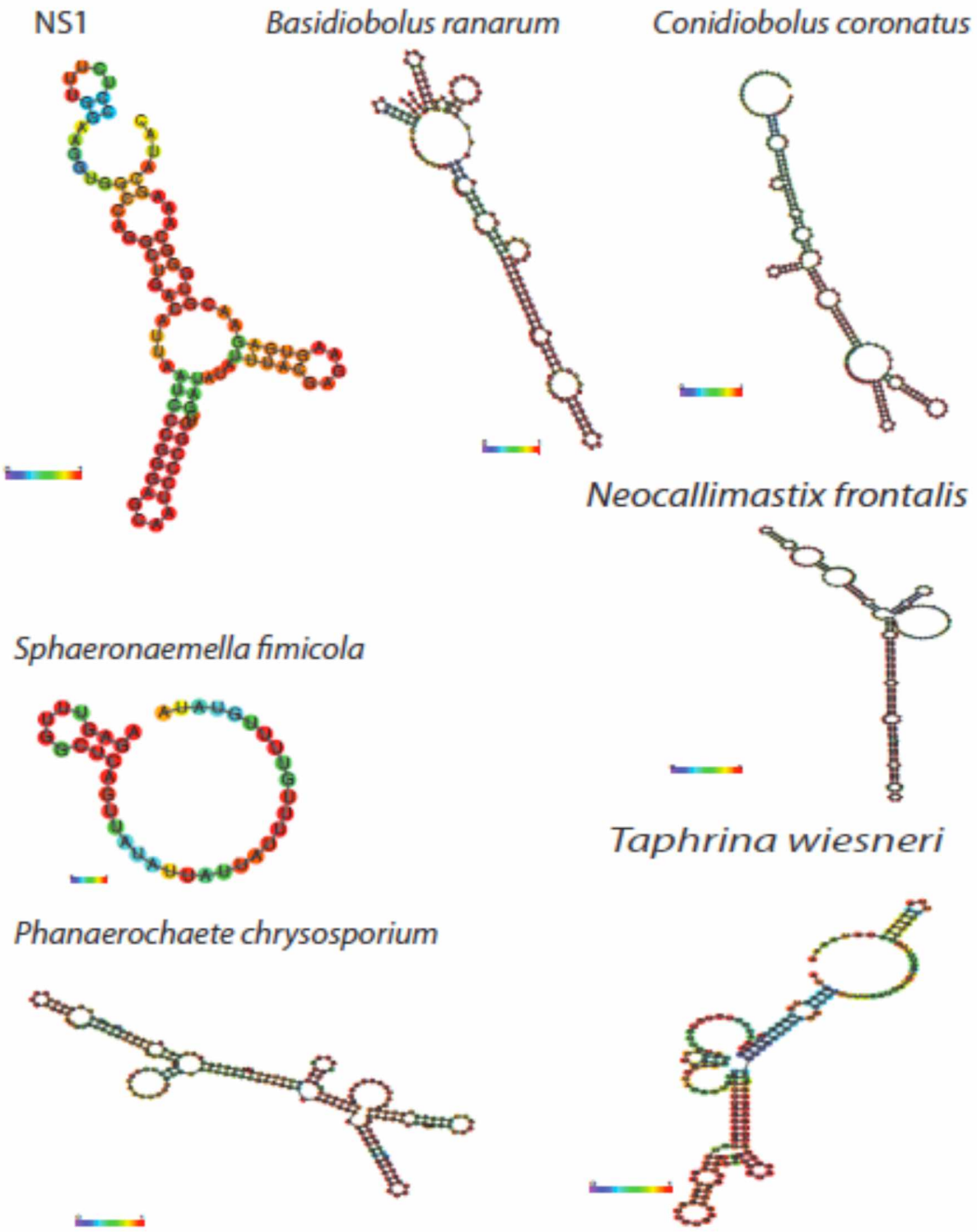


Figure 1.3. ITS1 secondary structure for NS1 as well as representative taxa belonging to most of the described fungal phyla. Colors depict positional probabilities. The red end of the spectrum indicates that the majority of modeled foldings (not only the MFE folding shown) support such a pairing while the blue end of the spectrum indicates little support for the pairing shown.

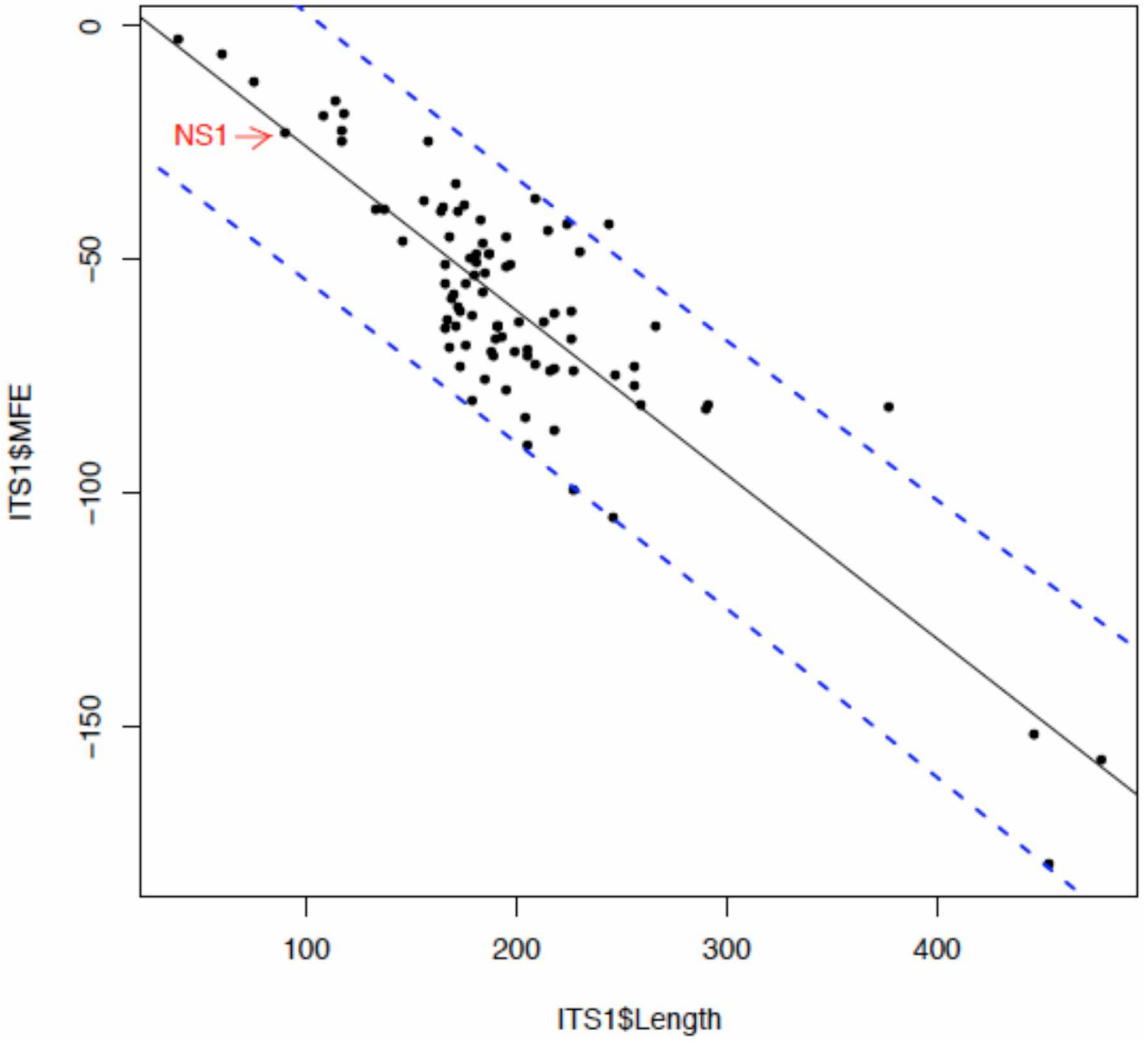


Figure 1.4. Linear regression of ITS1 length vs. MFE. Solid line is the trend line. Dotted lines depict the 95% confidence interval.

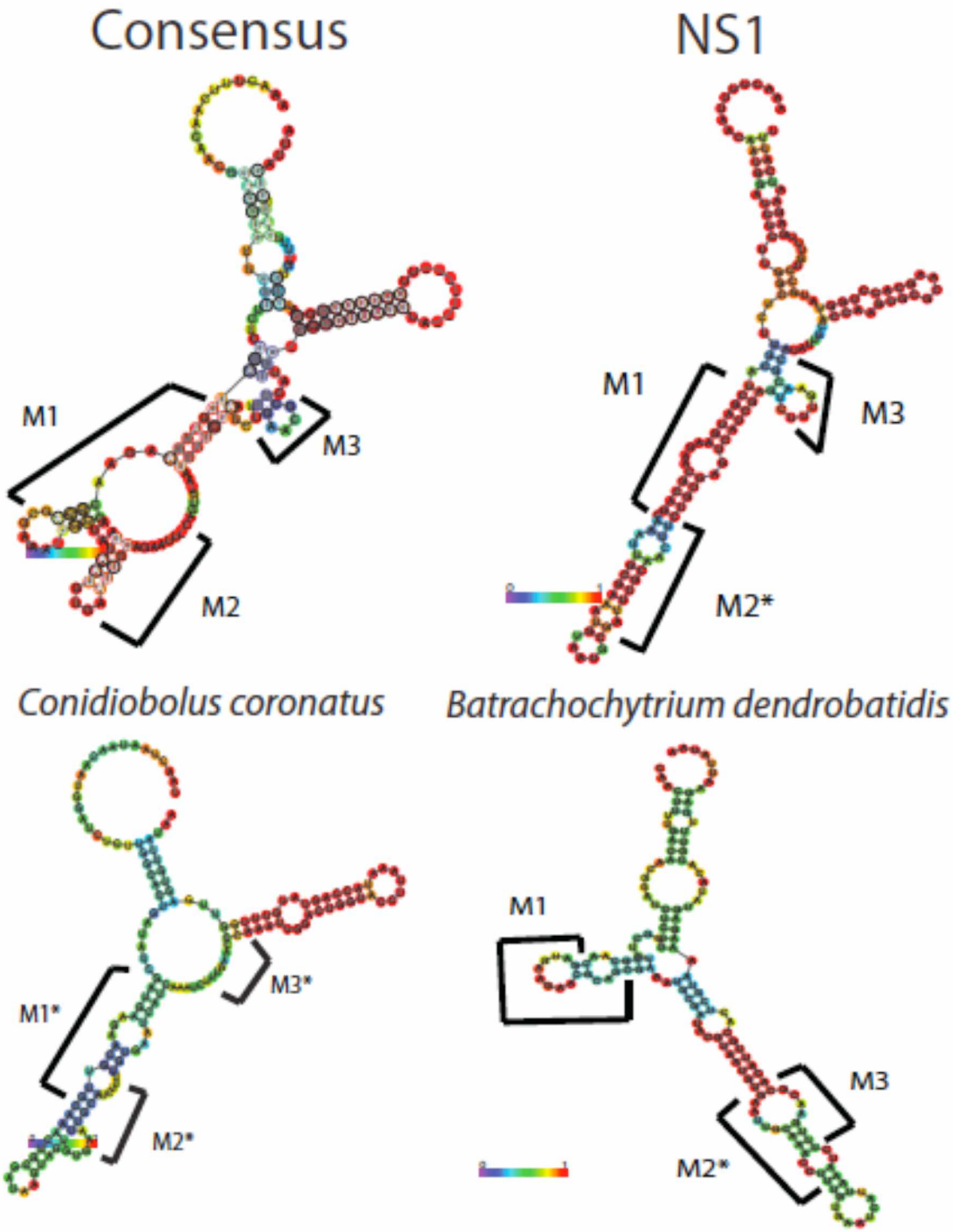


Figure 1.5. 5.8S secondary structures for NS1, two comparative fungal taxa belonging to the BFL, and a consensus folding of 188 fungi. Motifs one, two, and three are abbreviated M1-M3. Asterisks indicate a variation in sequence from that described in literature.

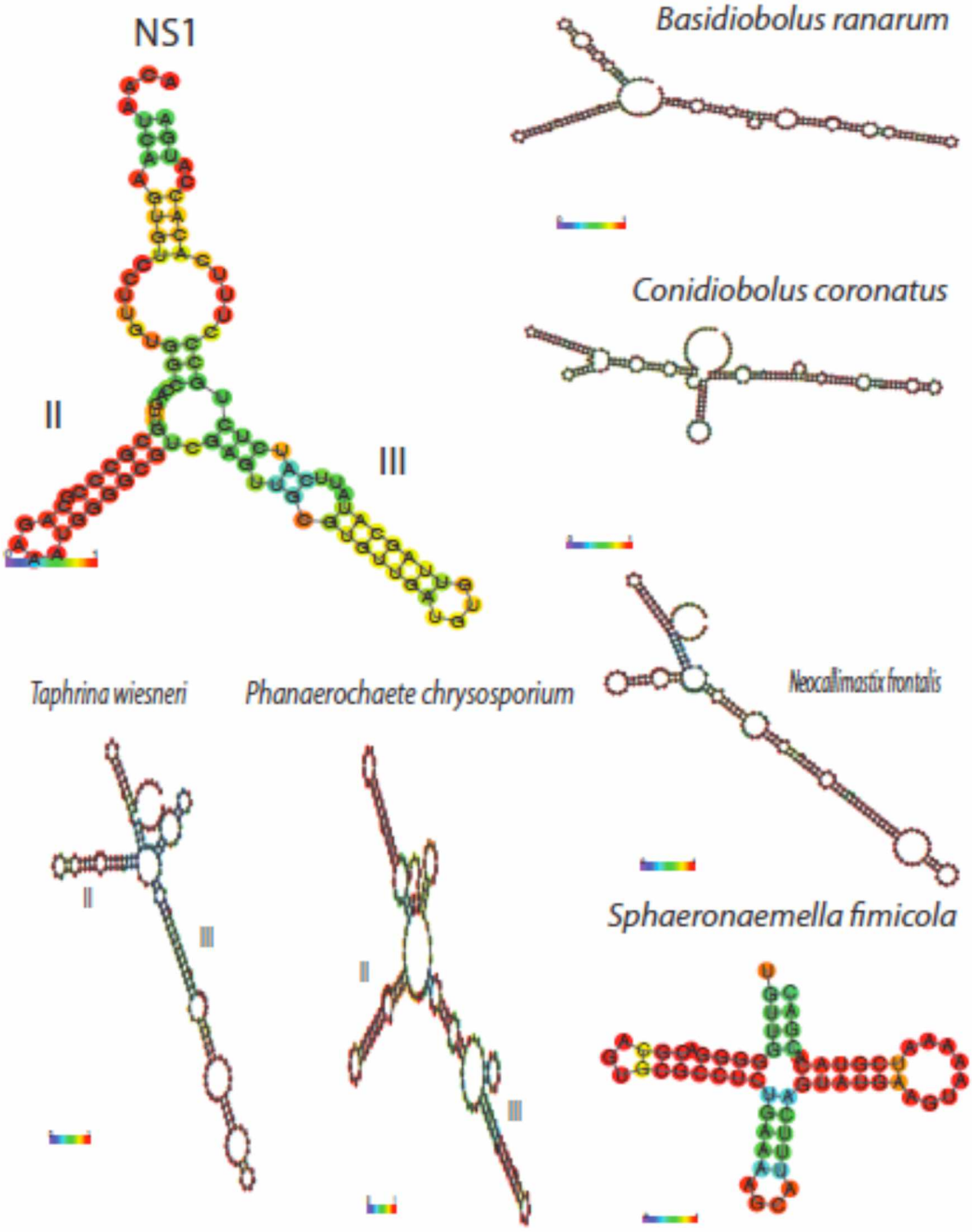


Figure 1.6. ITS2 secondary structure for NS1 as well as representative taxa belonging to most of the described fungal phyla. Domains one and two are numbered I and II respectively.

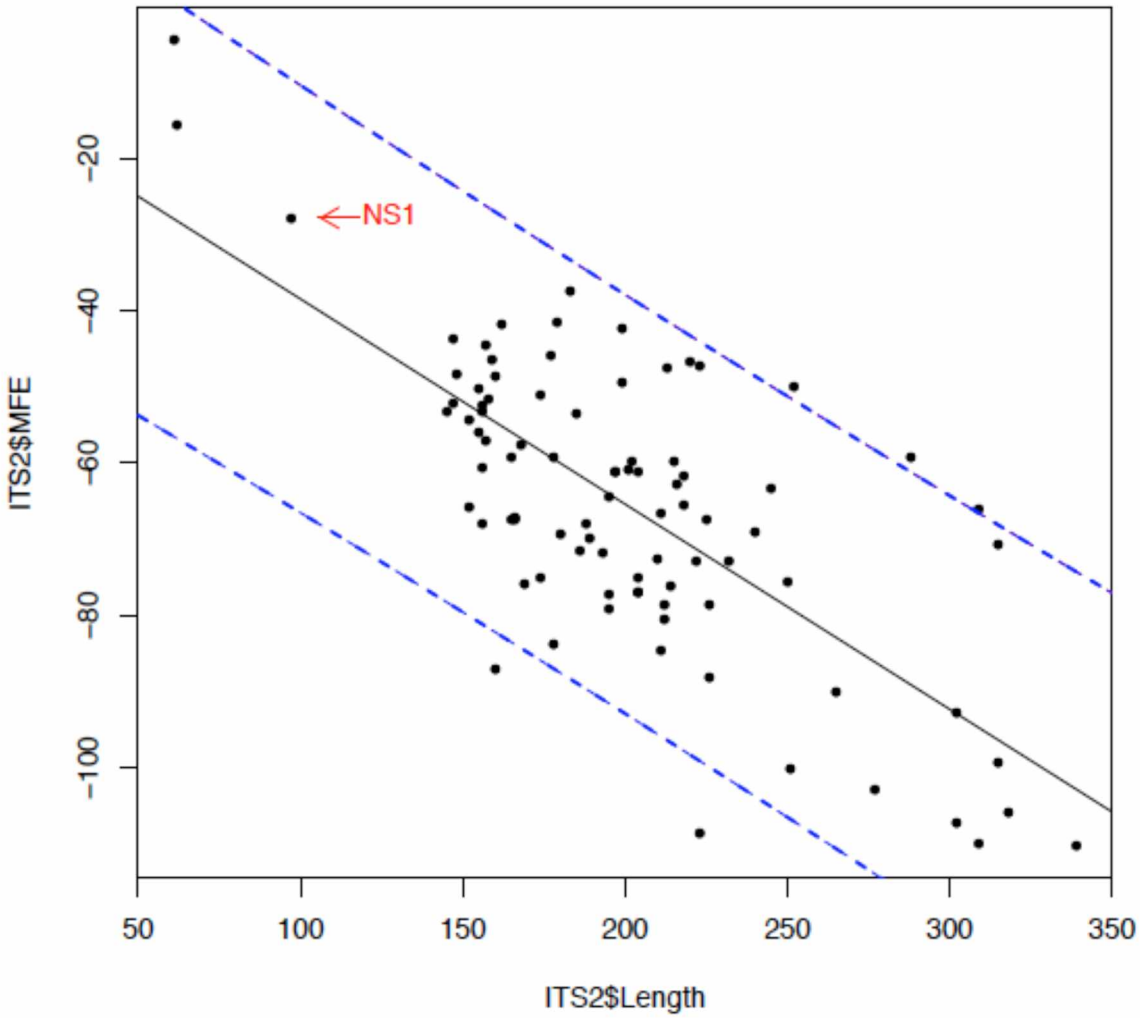


Figure 1.7. Linear regression of ITS2 length vs. MFE. Solid line is the trend line. Dotted lines depict the 95% confidence interval.

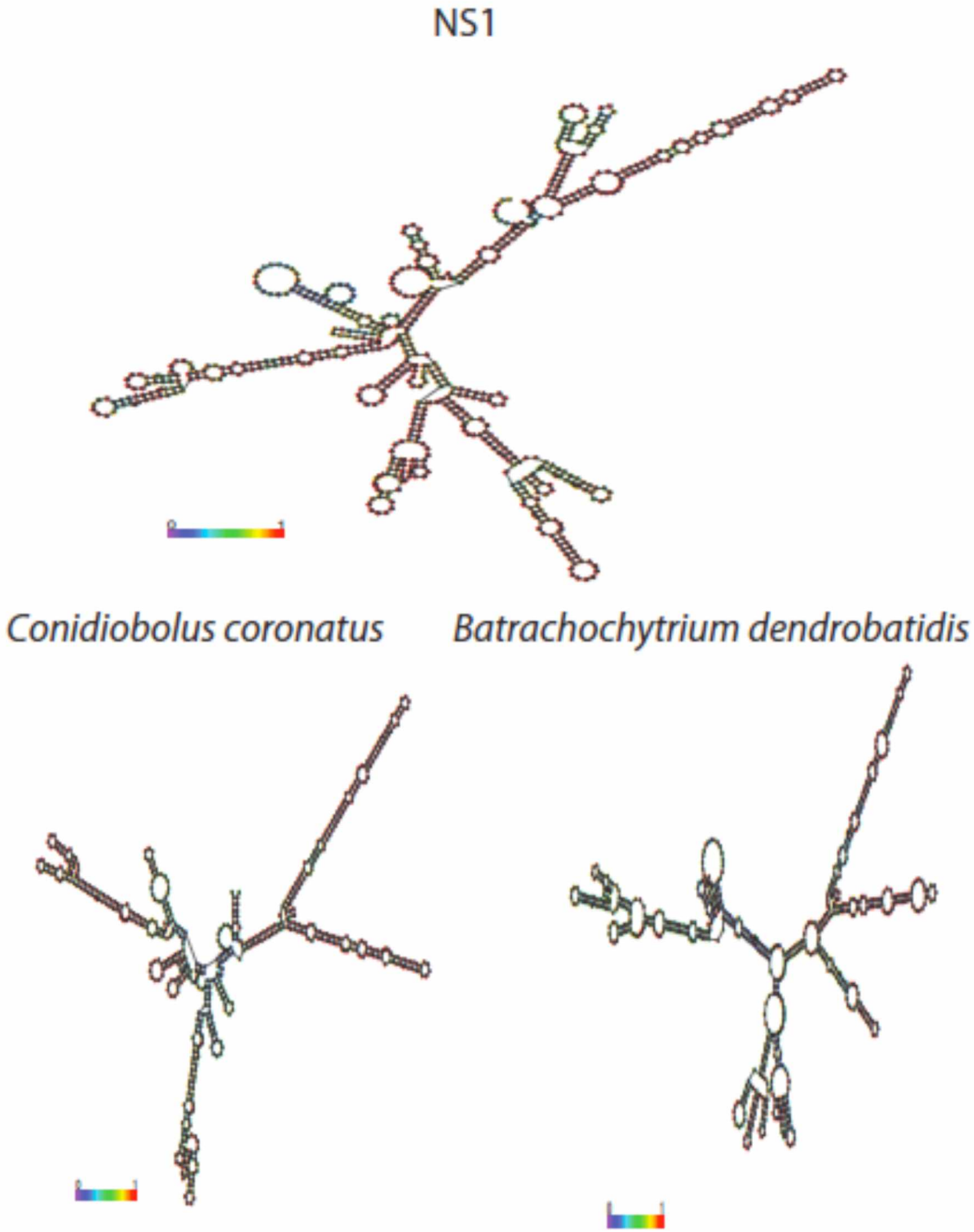


Figure 1.8. LSU secondary structure for NS1 and two members of the BFL.

Chapter 2:

Habitat preferences, micro-spatial structure, and temporal persistence of a novel fungal taxon in Alaskan boreal forest soils[1]

2.1 Abstract

A large-scale project to characterize boreal forest soil fungal community composition unearthed a particularly novel fungal sequence (NS1) that did not fall within known fungal phyla.

Through molecular methods alone, particularly the use of taxon-specific primers targeting the ITS gene-regions, I elucidated the habitat preferences, horizon preferences, and fine-scale spatial structure of NS1. I determined that it was associated with spruce, shows little interannual variability, and does not show a significant preference for either the organic or mineral horizon. It appears to have a clumped distribution. I also found little genetic variation within this putative taxon. None of the variability appeared to be associated with ecological factors and the variability was most likely a lab artifact. This lack of genetic variation may be because of primers that were too specific. This study has provided preliminary insights into the ecology of NS1 necessary before more in-depth studies can be pursued.

2.2 Introduction

Fungi fulfill many crucial ecological functions as decomposers and plant symbionts (O'Brien *et al.*, 2005). They likely play a particularly prominent role in boreal forests due to their ability to function at low temperatures, low pH, and in nutrient poor environments (Taylor *et al.*, 2010). Nevertheless, until relatively

---

[1] Glass,D.J., I.C. Herriott, N. Takebayashi, D.L. Taylor. Prepared for submission to Fungal Ecology

recently our understanding of boreal forest microbial ecological processes has largely been focused on functionality with little regard given to which species were fulfilling these functions (Taylor *et al*., 2010).

Beginning in 2003, a large-scale project to characterize soil fungal communities in the major stages of boreal forest succession using molecular methods was undertaken at the Bonanza Creek Long Term Ecological Research Site (BNZ-LTER) near Fairbanks, Alaska, USA (Taylor *et al*., 2010). Tens of thousands of soil clone sequences were analyzed and the majority of sequences appeared to be representatives of previously described fungal species, genera or families (D.L. Taylor, personal communication). Of those that did not fall into known fungal clades, one sequence in particular, hereafter referred to as novel sequence one (NS1), appeared especially unusual. Both BLAST (Altschul *et al.,* 1999) and phylogenetic analyses placed the sequence within the kingdom Eumycota (Chapter 1). If it is indeed not an artifact, the phylogenetic distance of NS1 from known fungal phyla suggests the discovery of a major fungal phylum previously unknown to humanity (Chapter 1).

I conducted several analyses of the inferred nuclear ribosomal RNA (rRNA) secondary structure of NS1 to evaluate if it was an authentic ribosomal gene-copy (Chapter 1). These included modeling the rRNA structure of the NS1 sequence and performing several analyses comparing the rRNA secondary structure of NS1 to other fungi to determine if conserved motifs, domains, and compensatory base changes were present. Most of these results were consistent with NS1 representing functional rRNA rather than a biological or lab artifact (Chapter 1). The original NS1 sequence was found in a soil clone library originating from a riparian black spruce (*Picea mariana*) stand at the Bonanza Creek LTER (LTER site code FP5C). A highly similar ~1200 bp sequence varying at only eight sites was found in another clone library from an upland site (UP2A) amplified using different primers. This further supports the authenticity of the sequence and suggests it may be found in multiple boreal forest habitats at the Bonanza Creek LTER.

Other highly novel lineages of fungi that have recently been revealed using molecular techniques have been shown to have multicontinental distributions (Porter *et al.*, 2008; Jones *et al.,* 2011; Rosling *et al.,* 2011). To date, NS1 is only known to occur in the Bonanza Creek LTER and at present can only be characterized using molecular techniques. Having already sought to determine the phylogenetic affinities of this putative novel taxon (Chapter 1) I will now delve into five major lines of inquiry regarding the ecology of this taxon: 1) its micro-spatial distribution, 2) its habitat preferences, 3) its soil horizon preferences, 4) its temporal persistence, 5) and what, if any, genetic variability is present within this taxon.

2.3 Methods

Initial Detection of NS1

Soil cores were collected from several sites within the Bonanza Creek LTER near Fairbanks, AK between 2003 and 2005. Large-scale PCR, cloning, and clone library sequencing were conducted to characterize the fungal communities present. Site descriptions, DNA extraction and amplification methods have been described previously (Geml *et al.*, 2010; Taylor *et al.*, 2010; Chapter One). In brief, the gene-region (~1200 bp) encompassing the ribosomal internal transcribed spacers (ITS) and a portion (~700 bp) of the ribosomal large subunit (LSU) was amplified from soil extracts using the fungal-specific PCR primers ITS1-F (Gardes and Bruns, 1993) and TW13 (Taylor and Bruns, 1999). Amplicons were cloned into pCR®4-TOPO vectors using a TOPO-TA kit (Invitrogen, Carlsbad, CA, USA) and sent to the Broad Institute of MIT and Harvard where transformations, automated clone-picking, and sequencing of clone libraries took place.

Micro-Spatial Structure

I utilized DNA extracts collected in a spatially explicit design to characterize the spatial distribution of NS1 across the LTER site where it was first detected. A total of 81 soil cores were collected from nine 3 m^2 plots in a stratified-random distribution across the 200 m by 200 m LTER site FP5C (Fig 2.1). Each of the nine plots was divided into nine 1 m^2 cells.

Soil core locations within each cell were chosen using a random number table. One soil core was taken from each cell. Cores were approximately 18 mm in diameter and 200 mm deep. The organic, humic, and mineral horizons were identified. Approximately 0.25 g of soil was subsampled from each of these horizons, frozen at -80 °C, and lyophilized. DNA extractions were conducted using an UltraClean-htp 96 Well Soil DNA kit (MoBio Laboratories Inc, Carlsbad, CA, USA). I then used the taxon-specific primers L2F1 (5'CCCGGTCGATATATTTACGAGAAG 3') and L2R2 (5'GGGCAGAGATGAATATGCTAACAC 3'; Chapter One) to target a 290 bp marker region for NS1 in DNAs from the organic horizon, i.e. the horizon from which the original NS1 clone originated. Illustra PureTaq Ready-To-Go PCR beads (GE Healthcare, Buckinghamshire, UK) and an MJ Research PTC-225 Peltier Thermalcycler (Harlow Scientific, Arlington, MA, USA) were used in all PCRs. Additionally, a negative control was included in each PCR setup. The following thermalcycler program was used: 96 °C for 2 min, followed by 35 cycles 94 °C for 30 s, 55 °C for 40 s, 72 °C for 1 min, and a final elongation step of 72 °C for 10 min. 5 μL of the reaction mixture was loaded into a 1.5% agarose gel, ran at 85V for 40 minutes on an Owl B1A EasyCast Mini Gel (Gel Size: (W x L) 7 x 8 cm; Footprint: (W x L x H) 10.5 x 16 x 9.5 cm; Running Buffer Volume: 400 ml; Owl Separation Systems, Inc. Portsmouth, NH), and visualized using EtBr. I was able to detect the presence of NS1 in three soil extracts using this method. Nevertheless, it should be noted that this method cannot confirm the absence of NS1. For each successful amplification I attempted to sequence the amplicon.

Big Dye Terminator v3.1 (Applied Biosystems, CA, USA) chemistry was used for direct cycle sequencing. I used Sephadex in Centri-Sep columns (Princeton Seperations Inc., NJ, USA) to purify BigDye Terminator reactions. These products were sequenced on an ABI 3130xl Genetic Analyzer (Life Technologies Corporation, Carlsbad, California, USA) at the University of Alaska Fairbanks Nucleic Acids Core Lab.

Habitat Preference

In order to assess the occurrence of NS1 across major forest and habitat types within Bonanza Creek, I applied the same PCR-screening to an array of soil DNA extracts collected in 2004 and 2005. The soil sampling, DNA extraction and analyses of total fungal communities have been described in Taylor *et al.* (2010). In brief, nine upland sites representing three stages of forest succession (early, mid and late succession) were screened. Fifty soil cores were collected at 10 m intervals along four parallel transects across each site. For each core, the organic and mineral soil horizons were separated and 0.25 g subsamples collected (Taylor *et al.,* 2010). In contrast to the samples from site FP5C, subsamples were pooled by horizon before being frozen at -80 °C and lyophilized. Twelve lowland black spruce sites representing a range of pH and moisture contents (described in Hollingsworth *et al*., 2006) were screened using these same methods. PCR was conducted as described above. Successful amplicons were sent to McLab (South San Francisco, CA, USA) for PCR cleanup and sequencing.

Soil Horizon Preference

Several prior studies have shown that fungal community composition varies across the soil vertical profile (Taylor and Bruns, 1999; Dickie *et al*., 2002), so I sought to determine if NS1 was predominately associated with either the organic or

mineral soil horizon by conducting Fisher's exact tests in R (n=30; Table 1; R Development Core Team, 2008). The mineral horizons of the nine upland sites collected in 2005 were screened for the presence of NS1 in same manner already described.

Temporal Persistence

In order to evaluate the temporal persistence of NS1 across multiple years on a fine spatial scale within a single site I collected nine soil cores from FP5C in 2009. These soil cores were collected from roughly the same locations as cores that revealed the presence of novel taxa in the 2003 samples. Soil extractions from 2009 samples were conducted using a PowerMax Soil DNA Isolation kit (MoBio Laboratories, Inc., Carlsbad, CA, USA). Approximately 1 gram of organic soil or 2 grams of mineral soil was used per extraction. DNA was amplified using the primers L2F1 and L2R2 to screen for NS1 as above. PCR screening was carried out as above and amplicons sent to McLab (South San Francisco, CA, USA) for PCR cleanup and sequencing.

Genetic Variation

The possibility that NS1 could represent a lineage comprising diverse species, as opposed to a monotypic taxon, led me to sequence amplicons and examine levels of genetic variation among samples. My 290 bp marker encompassed portions of the hypervariable ITS regions and was therefore likely to possess polymorphisms if multiple distinct representatives of the NS1 taxon were present. We compared sequences from different sites, soil horizons, and years in order to assess whether particular genetic variants were associated with particular environments or time-points.

Using Aligner (v. 1.5.2 CodonCode Corp., Dedham, MA), sequences were analyzed for single nucleotide polymorphisms (SNPs). When a site with a base that differed from the consensus base call was found I checked if a SNP was present. If at least one read direction had a phred score above 20 and the base appeared to be represented by a strong single peak, this position was designated a SNP. If neither read direction had a base call above 20 but both were between 10 and 20 and appeared to be represented by a dominant peak as opposed to sequencing error, it was designated a SNP. If one read direction was of low quality and the other was of high quality and did not appear to be represented by a SNP the base was not designated a SNP. If two or more dominant peaks appeared to be present at a single position, IUPAC ambiguity codes were used.

PAUP* v. 4.0b10 (Swofford, 2003) was used to conduct an exhaustive parsimony search and construct a phylogram. The input alignment was based on single nucleotide polymorphisms (SNPs) found within seven sequences originating from this study (Table 2). The original two soil clone sequences were used in this phylogenetic reconstruction as well. These sequences originated from soils representing every year of sample collection, the majority of habitats and both the organic and mineral soil horizons (Table 2). No outgroups were used because I was only interested in the relationship of NS1 variants to one another and in determining if soils that shared certain ecological factors, such as those analyzed above, shared more similar NS1 variants than did different soils.

2.4 Results/Discussion

Micro-Spatial Structure

I found NS1 in three out of 81 soil cores collected at site FP5C (Fig 2.1). Two of these soil cores were in the plot located in the northwest corner of the site. Prior studies of both sporocarps and soil extracts have shown certain fungal taxa to have

clumped distributions (Bergemann and Miller, 2002; Taylor and Bruns, 1999; Lilleskov *et al.,* 2004). This could be due to microsite preferences or to the expansion of an initial colonizing genet (Taylor and Bruns, 1999). Spatial clumping has been most often investigated in dominant taxa. It is noteworthy that this taxon appears to be extremely low in abundance (1 out of 384 passing clones in 2003 FP5C organic clone library; 1 out of 931 passing clones in 2004 UP2A organic clone library), yet also displays a clumped distribution. This result adds to the growing body of evidence demonstrating strong patchiness of individual fungi in soil at multiple spatial scales (Lilleskov *et al.*, 2004).

Habitat Preference

Although first detected in a riparian black spruce stand, NS1 was also found in all the upland sites that had a large proportion of white spruce present. These included all of the mature white spruce stands, and two mid-succession stands composed of nearly equal proportions of white spruce and birch. NS1 was not found in site UP2C, a mid-succession stand that is dominated by aspen, with a much lower presence of white spruce and paper birch. Nor was NS1 found in any UP1 sites. These are early succession stands dominated by a mixture of aspen, willow, or alder.

Surprisingly, although found repeatedly in the riparian black spruce stand, NS1 was not found in any other black spruce sites. The non-riparian lowland black spruce sites spanned a spectrum from dry and acidic to wet and non-acidic (Hollingsworth *et al.,* 2006). This suggests that other environmental factors besides dominant tree species may have a significant impact on the distribution of NS1. At the same time, NS1 was found in all of the sites that had a substantial white spruce abundance. Hence, it appears that both edaphic factors and overstory tree species may influence the distribution of NS1, but more research is needed.

Soil Horizon Preference

Molecular studies have shown a high degree of niche differentiation based on soil layers and major shifts in fungal community composition corresponding to soil horizons (Taylor and Bruns, 1999; Dickie *et al.*, 2002; Lindahl *et al.*, 2007). Dickie *et al.* (2002) found that although some fungi are multilayer generalists or litter layer generalists, many were constrained to a particular horizon. Overall, NS1 does not appear to have a significant soil horizon preference (Table 1). The organic horizon was sampled more extensively than the mineral horizon. NS1 was present in 11 of 120 organic soil extracts (Table 1). If the 90 non-pooled FP5C soil samples are excluded NS1 is found in six of 30 pooled organic soil extracts and only one of nine mineral extracts (Table 1). Fisher's exact test results indicate that there is no significant difference between soil horizons in regards to the presence of NS1 (n=30,df=1, p=0.36, Odds ratio=3.82). If confirmed by larger sampling efforts, particularly of the mineral horizon, this suggests that NS1 maybe a multilayer generalist.

Temporal Persistence

NS1 not only displayed a highly clumped spatial distribution in 2003, it largely adhered to this pattern when resampled in 2009 (Fig 2.1). Overall, NS1 appears to show little interannual variability between two sampling events spaced six years apart. Studies have shown that members of the Russulaceae that demonstrated a highly clumped distribution also demonstrated prolonged temporal persistence (Bergemann and Miller, 2002). In one case the same genets were found up to 11 years later (Bergemann and Miller, 2002). It has been hypothesized that taxa that display a highly clumped distribution that are also dominant members of the fungal community are successful because they initially colonize a location with bountiful resources, produce large clones, and are long lived (Taylor and Bruns,

1999). This hypothesis may not apply to NS1 because it is not a dominant taxon, nor is it likely to be ectomycorrhizal. Nevertheless, since it persisted in the same locations for six years, it does appear to be either long-lived or reproduces successfully on a local scale.

NS1 was also found in several upland LTER sites over multiple years. It was found in pooled soil extracts from the organic horizons of site UP2A in both 2004 and 2005, and pooled organic soil extracts from site UP3B in both 2004 and 2005. Just as NS1 shows temporal persistence at site FP5C on a very fine scale (within a few meters) it can be found in pooled samples at two other sites over multiple years as well.. Few studies have analyzed entire soil fungal communities across a site. Far fewer have compared soil fungal communities at the same site across multiple years (Izzo *et al.,* 2005). Izzo *et al.* (2005) found that the same dominant taxa were present on a coarse spatial scale (>25m) from year to year but community structure varied on a finer scale (5cm). Only 23% of species were found in the same plot every year (Izzo *et al.,* 2005). This high degree of community composition turnover in late successional forests makes the temporal persistence of NS1 particularly interesting.

Genetic Variation

Overall, I found little sequence variation among NS1 amplicons. Only 10 SNPs were detected, and many sequences had ambiguous base calls at these positions. I was not able to draw any ecological conclusions from the phylogram based upon them. If these polymorphisms do represent true biological variation between different representatives of the NS1 clade or intraindividual variation, more cloning and sequencing would be needed to confirm this.

This lack of genetic variation stands in stark contrast to the recently discovered clade of novel ascomycetes, soil clone group one (SCG1; Porter *et al.,* 2008). SCG1 was shown to be highly diverse, with many internal clades corresponding to the different sites where it was found. The phylogenetic placement

and high level of sequence variation found within this clade suggests it represents an ancient lineage that has since undergone a great deal of speciation (Porter *et al.,* 2008). Perhaps NS1 also represents a speciose lineage but my primers were too specific to detect closely related fungi and only amplified this particular taxon.

2.5 Conclusions

My simple PCR surveys demonstrate that it is possible to learn a substantial amount about the ecologies of species that are known only by their DNA. I have demonstrated that this deeply divergent, rare, novel fungus maybe a multilayer generalist, occurs in both upland and floodplain habitats that contain spruce, but appears to be very rare or absent from non-riparian lowlands. It also occurs in patches on the scale of several meters, and persists in sites over multiple years. I hope future efforts will expand our ecological understanding of this fungus. For example, fluorescent *in situ* hybridization (FISH) might allow us to visualize cells of the organism in environmental samples. This would help us ascertain the habitat preferences of these fungi in much finer detail and would shed light on their morphology and taxonomy.

2.6 Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, 1999. Basic local alignment search tool. *Journal of Molecular Biology* **215:** 403-410.

Bergemann SE and Miller SL, 2002. Size, distribution, and persistence of genets in local populations of the late-stage ectomycorrhizal basidiomycete, *Russula brevipes*. *New Phytologist* **156**: 313-320.

Dickie IA, Xu B, and Koide RT, 2002. Vertical niche differentiation of ectomycorrhizal hyphae in soil as shown by T-RFLP analysis. *New Phytologist* **156**:527-535.

Geml J., Laursen GA, Herriott IC , McFarland JM, Booth MG, Lennon N, Nusbaum HC, 2010. Phylogenetic and ecological analyses of soil and sporocarp DNA sequences reveal high diversity and strong habitat partitioning in the boreal ectomycorrhizal genus *Russula* (Russulales; Basidiomycota). *New Phytologist* **187**:494-507.

Gardes M, and Bruns TD, 1993. ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. *Molecular Ecology* **2**: 113-118.

Hollingsworth TN, Walker MD, Chapin FS, Parsons AL, 2006. Scale-dependent environmental controls over species composition in Alaskan black spruce communities. *Canadian Journal of Forest Research* **36**:1781-1796.

Izzo A, Agbowo J, and Bruns TD, 2005. Detection of plot-level changes in ectomycorrhizal communities across years in an old-growth mixed-conifer forest. *New Phytologist* **166**:619-630.

Jones MDM, Forn I, Gadelha C, Egan MJ, Bass D, Massana R, Richards TA, 2011. Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* **474:** 200-203.

Lilleskov EA, Bruns TD, Horton TR, Taylor DL, Grogan P, 2004. Detection of forest stand-level spatial structure in ectomycorrhizal fungal communities. *FEMS Microbiology Ecology* **49**:319-332.

Lindahl BD, Ihrmark K, Boberg J, Trumbore SE, Hogberg P, Stenlid J, Finlay RD, 2007. Spatial separation of litter decomposition and mycorrhizal nitrogen uptake in a boreal forest. *New Phytologist* **173**:611-620.

O'Brien HE, Parrent JL, Jackson JA, Moncalvo JM, Vilgalys R, 2005. Fungal community analysis by large-scale sequencing of environmental samples. *Applied and Environmental Microbiology* **71**:5544-5550.

Porter TM, Schadt CW, Rizvi L, Martin AP, Schmidt SK, Scott-Denton L, Vilgalys R, 2008. Widespread occurrence and phylogenetic placement of a soil clone group adds a prominent new branch to the fungal tree of life. *Molecular Phylogenetics and Evolution* **46**:635-644.

R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3 -900051-07-0, URL http://www.R-project.org.

Rosling A, Cox F, Cruz-Martinez K, Ihrmark K, Grelet G, Lindahl BD, Menkis A, James TY, 2011. Archaeorhizomycetes: Unearthing an ancient class of ubiquitous soil fungi. *Science* **333:**876-879.

Swofford DL, 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Taylor, DL, and Bruns TD, 1999. Community structure of ectomycorrhizal fungi in a *Pinus muricata* forest: minimal overlap between the mature forest and resistant propagule communities. *Molecular Ecology* **8**:1837-1850.

Taylor DL, Herriott IC, Stone KE, McFarland JW, Booth MG, Leigh MB, 2010. Structure and resilience of fungal communities in Alaskan boreal forest soils. Canadian Journal of Forest Research **40**:1288-1301.

2.7 Tables

Table 2.1. Presence of NS1 in soil-extract PCR survey of various LTER soil extracts. Black spruce is abbreviated as BS. Totals are given in bold.

Description	***NS1***	***Total***
Non-pooled soil extracts		
FP5C (BS) 2003 Organic	3	81
FP5C (BS) 2009 Organic	2	9
Pooled Soil Extracts		
Upland 2004 Organic	3	9
Upland 2005 Organic	3	9
TKN (BS) Organic	0	12
Organic Total	**11**	**120**
Upland 2005 Mineral	1	9
Total	**12**	**129**

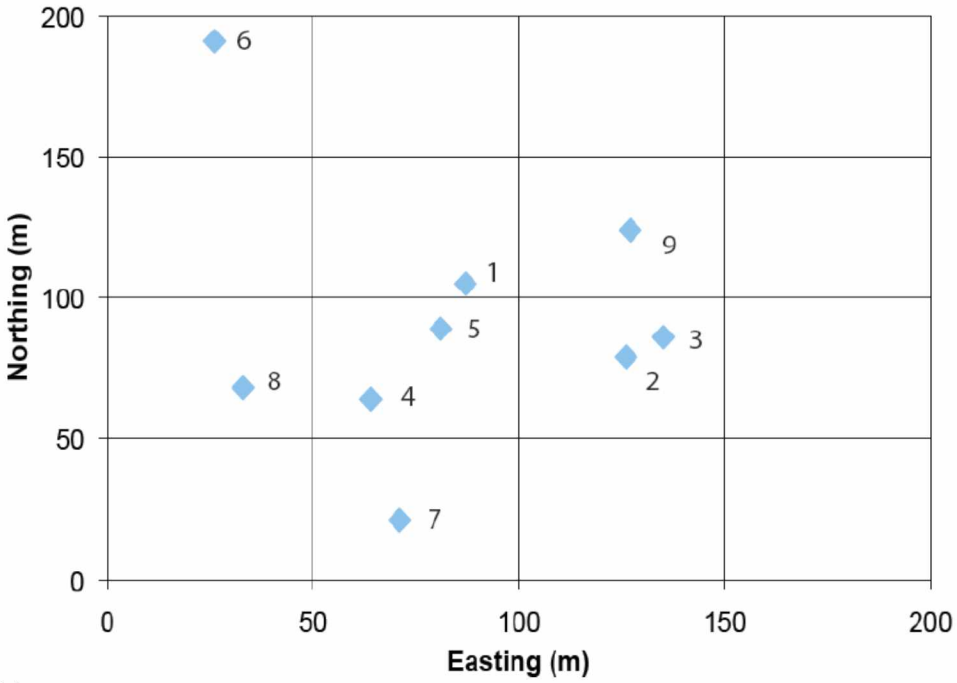
Table 2.2. Origins of NS1 sequence variants. A * indicates inclusion in SNP analyses.

Sites NS1 Present (Year; Horizon)	***Sequence Obtained***
UP2A (2004; Organic)	Yes*
UP2A (2005; Organic)	No
UP2B (2005; Mineral)	Yes*
UP3A (2004; Organic)	No
UP3B (2004; Organic)	Yes*
UP3B (2005; Organic)	Yes*
UP3C (2004; Organic)	Yes*
FP5C (2003; Organic) Plot 1 Cell 7	Yes
FP5C (2003; Organic) Plot 6 Cell 2	Yes
FP5C (2003: Organic) Plot 6 Cell 9	Yes
FP5C (2009; Organic) Plot 6 Cell 8	Yes*
FP5C (2009; Organic) Plot 6 Cell 9	Yes*

.

2.8 Figure

a)



b)

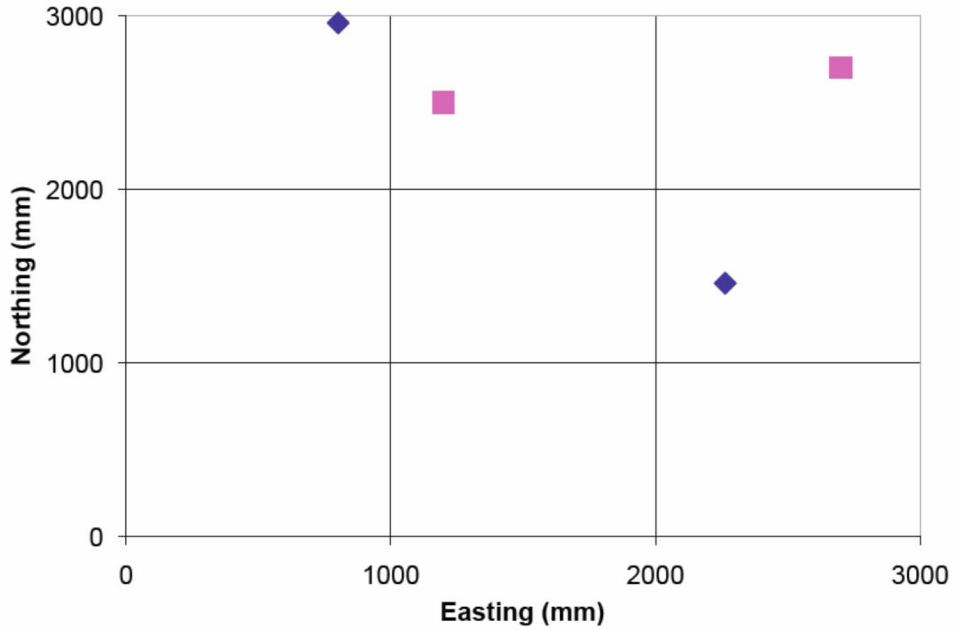


Figure 2.1. a) Distribution of NS1 across the riparian black spruce site FP5C. The 200m x 200m site is shown. Here, the distribution of the nine plots can also be seen. b) Plot six is expanded to show the distribution of NS1 in 2003 (blue) and 2009 (pink).

Conclusions

Solely through targeting DNA, I uncovered a wealth of information on the gene-authenticity and phylogenetic affinities of the putative novel fungal taxon NS1. Through the use of ribosomal RNA secondary structure modeling analyses I determined that NS1 rRNA secondary structure is consistent with the trends found in other fungi. My results that were not consistent with patterns described in the literature appeared to differ in other fungi, particularly those belonging to the basal fungal lineages, as well. Together my findings suggest that NS1 represents an authentic gene-copy and not a biological or lab artifact. My phylogenetic results suggest that this novel sequence is fungal in origin and more than likely belongs within the basal fungal lineages. Together with my rRNA secondary structure results this strongly suggests that NS1 represents a highly novel fungal taxon and potentially could represent a new class or phylum of fungi. Amplification of longer gene-regions could further resolve and increase support for the phylogenetic placement of NS1 in the fungal kingdom.

I also derived basic insights into the ecology of this rare fungal taxon without any knowledge of the organism besides knowing how to selectively target its DNA. These include descriptions of the spatial structure, landscape distribution, and temporal persistence of this novel fungal taxon across the Bonanza Creek LTER. This putative fungus appears to be a multilayer generalist. Furthermore, it occurs in both upland and riparian habitats that contain either black or white spruce but does not appear to be present in non-riparian lowlands, even if black spruce is the dominant tree type. Where NS1 is present, it appears to have a very clumped distribution on a sub-meter scale. It also appears to persist over long periods of time, being found in roughly the same locations six years apart. This suggests that even though it is a rare taxon, it is either long-lived or reproductively successful on a very local scale. Together this suggests that both soil type and overstory tree species have a strong impact on the distribution of NS1, but further research is needed into

the specific biotic and abiotic factors that impact its distribution. For example, through analyzing soils on a much finer scale (<1cm) it may be possible to determine if NS1 is associated with particular overstory trees, other plants, animals, or particular soil types.

Clearly, it is possible to learn a great deal about an organism just through the use of DNA-based methods. Although this approach may not be necessary for large and abundant organisms such as many plant and animal species it is frequently the only tool available to study many microbes. The use of molecular methods is becoming one of mycology's greatest tools for teasing out the remaining unknown fungal diversity. This approach is often combined with more traditional mycological methods to describe novel diversity once detected through molecular techniques. My study has shown that many basic insights into the systematics and ecology of organisms can be gleaned using molecular techniques alone. One of these, the validation of gene authenticity through the use of secondary structure modeling, is currently underemployed, and does not appear to have been used on environmental sequences to date. This technique can quickly validate whether a sequence is an artifact and should be more widely adopted, especially when molecular approaches are the only option for gaining insights into an organism's biology.

Literature Cited

Allen, T., T. Millar, S.M. Berch, and M.L. Berbee (2003). "Culturing and direct DNA extraction find different fungi from the same ericoid mycorrhizal roots." New Phytologist **160:** 255-272.

Balakirev, E. S. and F.J. Ayala (2003). "Pseudogenes: Are they "Junk" or functional DNA?" Annual Review of Genetics **37**: 123-151.

Blackwell, M. (2011). "The Fungi: 1, 2, 3 ... 5.1 million species?" American Journal of Botany **98**: 426-438.

Hawksworth, D.L. (2001). "The magnitude of fungal diversity: the 1.5 million species estimate revisited." Mycological Research **105**: 1422-1432.

Lilleskov, E.A., T.J. Fahey, T.R. Horton, and G.M. Lovett (2002). "Belowground ectomycorrhizal fungal community change over a nitrogen deposition gradient in Alaska." Ecology **83**: 104-115.

Pace, N.R. (1997). "A Molecular View of Microbial Diversity and the Biosphere." Science **276**: 734-740.

Schadt, C.W., A.P. Martin, D.A. Lipson, and S.K. Schmidt (2003). "Seasonal dynamics of previously unknown fungal lineages in tundra soils." Science **301**: 1359-1361.

Taylor, D.L. and T.D. Bruns (1999). "Community structure of ectomycorrhizal fungi in a *Pinus muricata* forest: minimal overlap between the mature forest and resistant propagule communities." Molecular Ecology **8**: 1837-1850.

Taylor, D.L., I.C. Herriott, K.E. Stone, J.W. McFarland, M.G. Booth, and M.B. Leigh (2010). "Structure and resilience of fungal communities in Alaskan boreal forest soils." Canadian Journal of Forest Research **40**: 1288-1301.

Vilgalys, R., and M. Hester (1990)." Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several Cryptococcus species." Journal of bacteriology **172**:4238-4246.

White, T.J., T. Bruns, S. Lee, and J.W. Taylor (1990). "Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics." Pp. 315-322 In: PCR Protocols: A Guide to Methods and Applications, eds. Innis, M. A., D. H. Gelfand, J. J. Sninsky, and T. J. White. Academic Press, Inc., New York. ©W°

Appendix

Four other highly novel sequences (NS2-NS5) were found in this study. Two (NS3 and NS4) appear to belong to the recently described clade of ascomycetes deemed soil clone group one (SCG1). The others appear to fall within the basal fungal lineages. Phylogenies containing these taxa, are shown. A table showing primers designed in this study to amplify longer gene-regions is shown as well.

Table A.1. Primers designed to target novel taxa (NS1-NS5). All primer sequences are oriented 5' to 3'. Lab results using these primers are shown as well. Wrong size is abbreviated as WS. I designed taxon specific ITS primers to target a ~300 bp region of ITS. I also designed primers targeting a taxon-specific region of the ITS to be paired with previously published LSU and SSU primers. These include the TW13 (White *et al.,* 1990), nLSU1221R (Schadt *et al.,* 2003), and TWI4 (Taylor and Bruns, 1999) primers to target the LSU and the NS1 (White *et al.,* 1990), SR1R (Vilgalys and Hester, 1990), and NS3 (White *et al.,* 1990) primers to target the SSU.

ITS

Target Taxon	Primer Name	Target Length (bp)	Primer Sequence	Amplicon	Cloned	Sequenced	Identity
NS1	L2F1	290	CCCGGTCGATATATTTACGAGAAG	Single	Yes	Yes	Match original
	L2R2		GGGCAGAGATGAATATGCTAACAC				
NS2	L4F1	277	TACAATTTTACCAGCGTCAAGCAC	Multiple	Yes	Yes	Botryotinia fuckeliana hypothethical protein
	L4R1		CCTCATCTTACTCGCCCTAGAGTC				
NS3	J22F1	293	GGGATCATTAAAAAGATGGGTCCT	Single	Yes	Yes	Match original
	J22R1		ATGGTTTTAAATGACGCTCAGACA				
NS4	B23F1	215	CCGCTTTTATGCCTTAGTCAGTCT	Single	Yes	Yes	Match original
	B23R1		ATGGTTTTAAATGACGCTCAGACA				
NS5	A8F1	300	AGGGCTTAAACATGGTGTTGAAAT	Single	Yes	Yes	Match original
	A8R1		TCCACAATCTTTAACCAACAAAGG				

LSU

Target Taxon	Primer Name	Target Length (bp)	Primer Sequence	Amplicon	Cloned	Sequenced	Identity
NS1	L2F1	~1000	CCCGGTCGATATATTTACGAGAAG	Single	Yes	Yes	Undetermined
	TW13		GGTCCGTGTTTCAAGACG				

Table A.1. continued

NS1	L2F1	~1500	CCCGGTCGATATATTTACGAGAAG
	nLSU1221R		CTAGATGAACYAACACCTT
NS1	L2F1	~1500	CCCGGTCGATATATTTACGAGAAG
	TW14		GCTATCCTGAGGGAAACTTC
NS1	L2LSUF1	~1300	GTCTTCCGATCAAGTGTGCTTTA
	TW14		GCTATCCTGAGGGAAACTTC
NS1	L2LSUF2	~1500	GTTAAAAGGGAAACGATTGGAAC
	TW14		GCTATCCTGAGGGAAACTTC
NS1	L2LSUF2	~1500	GTTAAAAGGGAAACGATTGGAAC
	TW14		GCTATCCTGAGGGAAACTTC
NS1	ITS149f	1400	ATTTACGAGAAGTGAGAACG
	nLSU1221R		CTAGATGAACYAACACCTT
NS1	ITS140f	1400	GGTCGATATATTTACGAGAA
	nLSU1221R		CTAGATGAACYAACACCTT
NS1	ITS140f	~2000	GGTCGATATATTTACGAGAA
	TW14		GCTATCCTGAGGGAAACTTC
NS1	ITS390f	~2000	GTTGCGTGTTGATGTGTTAG
	TW14		GCTATCCTGAGGGAAACTTC
NS1	ITS425f	~2000	CCTTTCACACCATGATCTC
	TW14		GCTATCCTGAGGGAAACTTC
NS2	L4F1	~1000	TACAATTTTACCAGCGTCAAGCAC
	TW13		GGTCCGTGTTTCAAGACG
NS2	L4F1	~1000	TACAATTTTACCAGCGTCAAGCAC
	TW13		GGTCCGTGTTTCAAGACG
NS2	L4F1	~2000	TACAATTTTACCAGCGTCAAGCAC
	TW14		GCTATCCTGAGGGAAACTTC
NS2	L4F1	~2000	TACAATTTTACCAGCGTCAAGCAC
	nLSU1221R		CTAGATGAACYAACACCTT
NS3	J22F1	~1000	GGGATCATTAAAAAGATGGGTCCT
	TW13		GGTCCGTGTTTCAAGACG
NS3	J22F1	>1500	GGGATCATTAAAAAGATGGGTCCT

None	NA	NA	NA
None	NA	NA	NA
WS	No	Yes	Undetermined
WS	No	No	NA
Single	Yes	Yes	*Chytridium* sp, *Mucor* sp
Multiple	Yes	No	NA
None	NA	NA	NA
None	NA	NA	NA
None	NA	NA	NA
None	NA	NA	NA
Single	Yes	No	NA
Single	Yes	No	No
None	NA	NA	NA
None	NA	NA	NA
Single	No	No	NA
Single	No	NA	NA

Table A.1. continued

	TW14		GCTATCCTGAGGGAAACTTC
NS3	J22F1	>1500	GGGATCATTAAAAAGATGGGTCCT
	nLSU1221R		CTAGATGAACYAACACCTT
NS4	B23F1	~1000	CCGCTTTTATGCCTTAGTCAGTCT
	TW13		GGTCCGTGTTTCAAGACG
NS4	B23F1	>1500	CCGCTTTTATGCCTTAGTCAGTCT
	TW14		GCTATCCTGAGGGAAACTTC
NS4	B23F1	>1500	CCGCTTTTATGCCTTAGTCAGTCT
	nLSU1221R		CTAGATGAACYAACACCTT
NS5	A8F1	~1000	AGGGCTTAAACATGGTGTTGAAAT
	TW13		GGTCCGTGTTTCAAGACG
NS5	A8F1	>1500	AGGGCTTAAACATGGTGTTGAAAT
	TW14		GCTATCCTGAGGGAAACTTC
NS5	A8F1	>1500	AGGGCTTAAACATGGTGTTGAAAT
	nLSU1221R		CTAGATGAACYAACACCTT

			SSU
Target Taxon	**Primer Name**	**Target Length (bp)**	**Primer Sequence**
NS1	NS1	>1500	GTAGTCATATGCTTGTCTC
	L2R2		GGGCAGAGATGAATATGCTAACAC
NS1	SR1R	>1500	TACCTGGTTGATTCTGCCAGT
	L2R2		GGGCAGAGATGAATATGCTAACAC
NS1	NS1	>1500	GTAGTCATATGCTTGTCTC
	L2SSUR1		CTTTGCCCACGTTCTCACTT
	NS1	>1500	GTAGTCATATGCTTGTCTC
	L2SSUR2		TATGCTTTGCCCACGTTCTC
NS1	NS1	>1500	GTAGTCATATGCTTGTCTC
	ITS130r		ACGTTCTCACTTCTCGTAAATA

Single	Yes	Yes	SCG1
Single	No	No	NA
Single	No	No	NA
Single	Yes	Yes	SCG1
Single	Yes	No	NA
None	NA	NA	NA
None	NA	NA	NA

Amplicon	**Cloned**	**Sequenced**	**Identity**
None	NA	NA	NA
None	NA	NA	NA
Single	No	Yes	*Alpova diplophloeus* SSU
Single	No	No	NA
Single WS	No	NA	NA

Table A.1. continued

NS1	SR1R	>1500	TACCTGGTTGATTCTGCCAGT
	ITS130r		ACGTTCTCACTTCTCGTAAATA
NS1	NS1	>1500	GTAGTCATATGCTTGTCTC
	ITS171r		CCACGTTCTCACTTCTCGTA
NS1	SR1R	>1500	TACCTGGTTGATTCTGCCAGT
	ITS171r		CCACGTTCTCACTTCTCGTA
NS2	NS1	>1500	GTAGTCATATGCTTGTCTC
	L4R1		CCTCATCTTACTCGCCCTAGAGTC
NS2	NS3	>1500	GCAAGTCTGGTGCCAGCAGCC
	L4R1		CCTCATCTTACTCGCCCTAGAGTC
NS2	SR1R	>1500	TACCTGGTTGATTCTGCCAGT
	L4R1		CCTCATCTTACTCGCCCTAGAGTC
NS3	NS1	>1500	GTAGTCATATGCTTGTCTC
	J22R1		ATGGTTTTAAATGACGCTCAGACA
NS3	NS1	>1500	GTAGTCATATGCTTGTCTC
	J22R1		ATGGTTTTAAATGACGCTCAGACA
NS3	SR1R	>1500	TACCTGGTTGATTCTGCCAGT
	J22R1		ATGGTTTTAAATGACGCTCAGACA
NS4	NS1	>1500	GTAGTCATATGCTTGTCTC
	B23R1		ATGGTTTTAAATGACGCTCAGACA
NS4	SR1R	>1500	TACCTGGTTGATTCTGCCAGT
	B23R1		ATGGTTTTAAATGACGCTCAGACA
NS5	NS1	>1500	GTAGTCATATGCTTGTCTC
	A8R1		TCCACAATCTTTAACCAACAAAGG
NS5	SR1R	>1500	TACCTGGTTGATTCTGCCAGT
	A8R1		TCCACAATCTTTAACCAACAAAGG

None	NA	NA	NA
None	NA	NA	NA
None	NA	NA	NA
Multiple WS	No	NA	NA
Multiple WS	No	NA	NA
Streaking	No	NA	NA
Streaking Multiple WS	No	NA	NA
Single	No	Yes	NA
Single	Yes	No	NA
Single	Yes	Yes	NA
Single	No	Yes	NA
Single	Yes	No	NA
Single	Yes	No	NA
None	NA	NA	NA

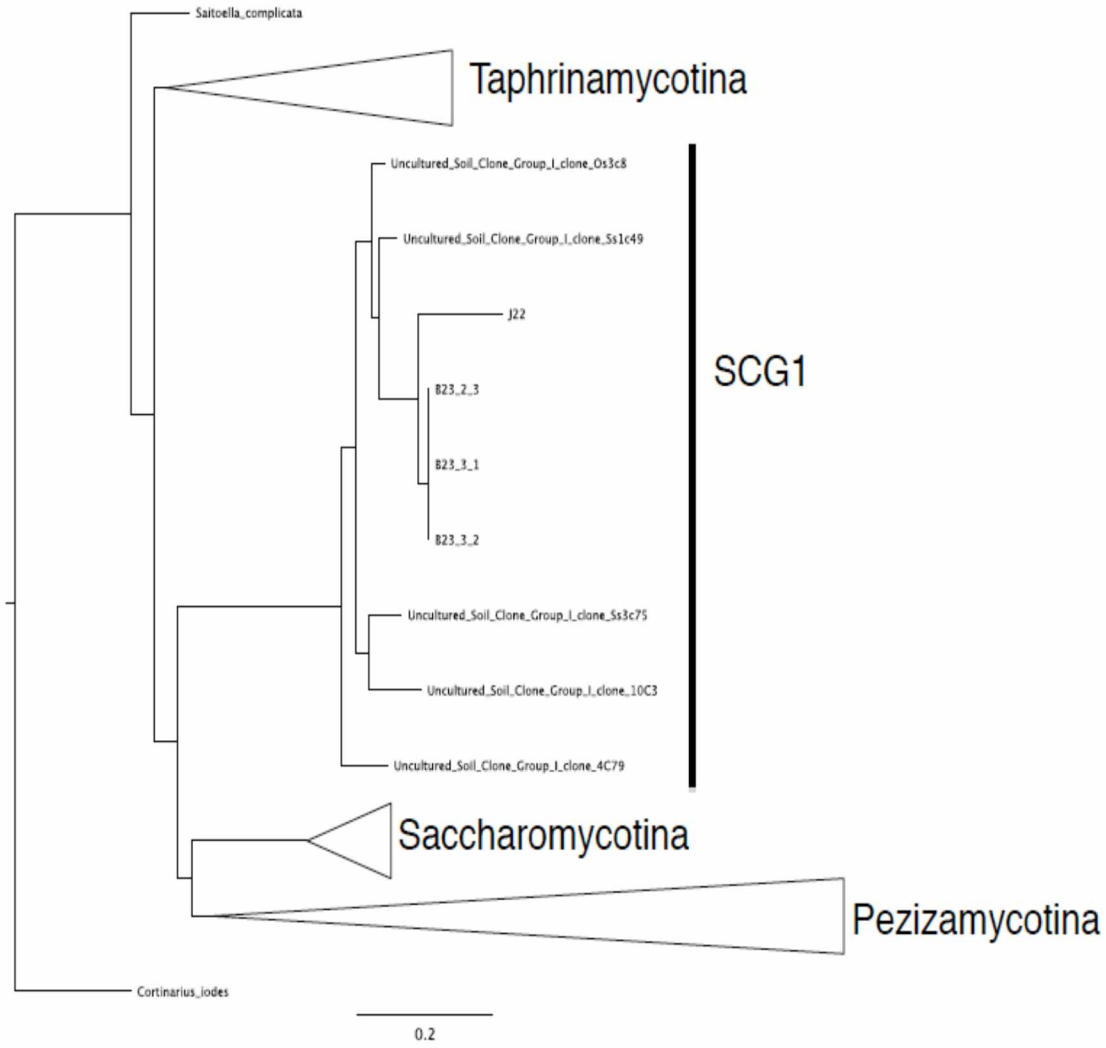


Figure A.1. Phylogeny of ascomycetes including novel taxa NS3 (J22) and NS4 (B23).

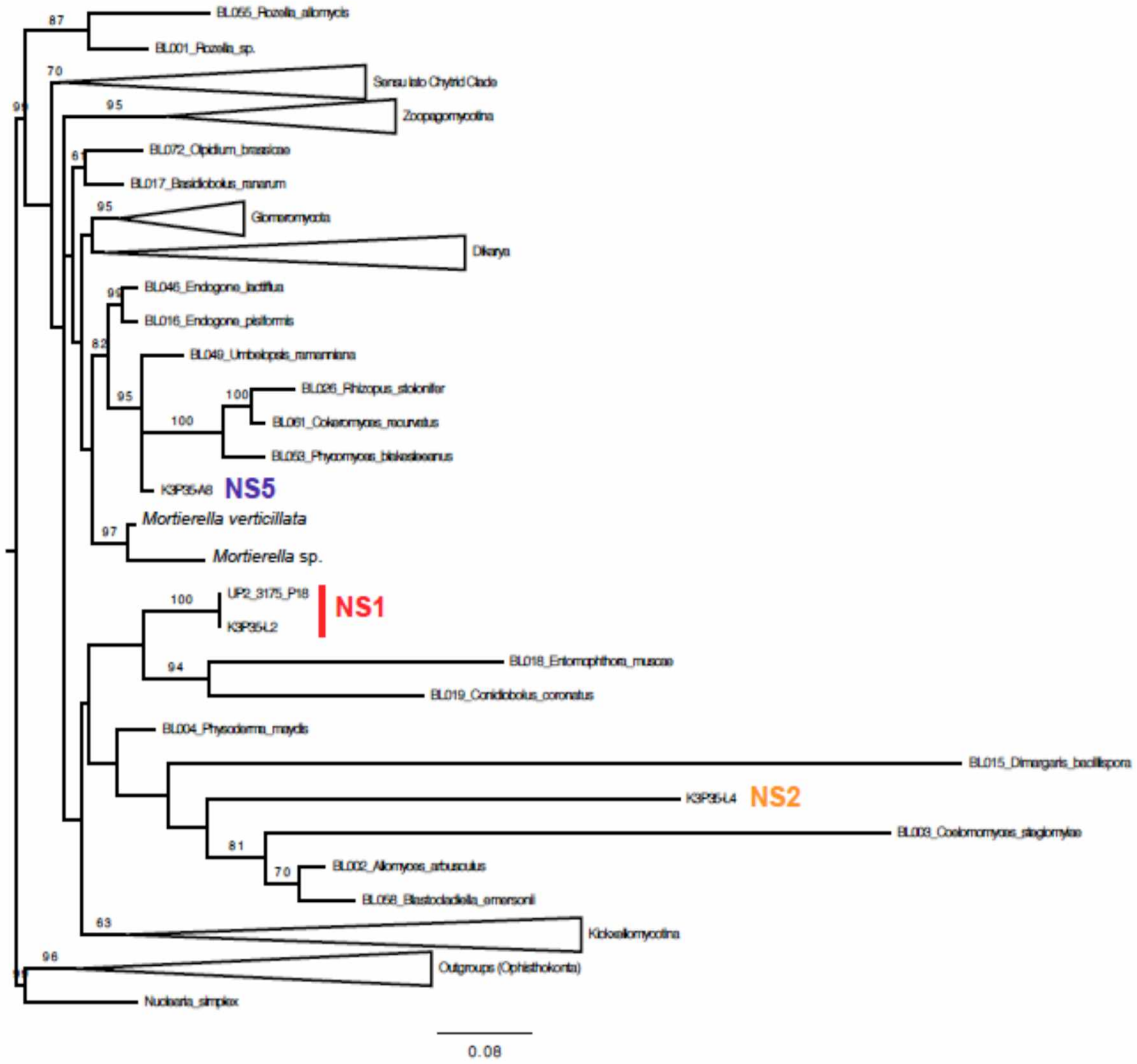


Figure A.2. Phylogeny of basal fungi including novel taxa NS1, NS2 and NS5.